

Philosophical Studies Series 117

Georgios Anagnostopoulos  
*Editor*

Socratic, Platonic  
and Aristotelian  
Studies: Essays  
in Honor of  
Gerasimos Santas

 Springer

Socratic, Platonic and Aristotelian Studies:  
Essays in Honor of Gerasimos Santas

# PHILOSOPHICAL STUDIES SERIES

---

VOLUME 117

---

*Founded by Wilfrid S. Sellars and Keith Lehrer*

*Editor*

Stephen Hetherington, *The University of New South Wales, Sydney, Australia*

*Senior Advisory Editor*

Keith Lehrer, *University of Arizona, Tucson, AZ, U.S.A.*

*Associate Editor*

Stewart Cohen, *University of Arizona, Tucson, AZ, U.S.A.*

*Board of Consulting Editors*

Lynne Rudder Baker, *University of Massachusetts, Amherst, MA, U.S.A.*

Radu Bogdan, *Tulane University, New Orleans, LA, U.S.A.*

Marian David, *University of Notre Dame, Notre Dame, IN, U.S.A.*

John M. Fischer, *University of California, Riverside, CA, U.S.A.*

Allan Gibbard, *University of Michigan, Ann Arbor, MI, U.S.A.*

Denise Meyerson, *Macquarie University, NSW, Australia*

François Recanati, *Institut Jean-Nicod, EHESS, Paris, France*

Mark Sainsbury, *University of Texas, Austin, TX, U.S.A.*

Stuart Silvers, *Clemson University, Clemson, SC, U.S.A.*

Barry Smith, *State University of New York, Buffalo, NY, U.S.A.*

Nicholas D. Smith, *Lewis & Clark College, Portland, OR, U.S.A.*

Linda Zagzebski, *University of Oklahoma, Norman, OK, U.S.A.*

For further volumes:

<http://www.springer.com/series/6459>

Georgios Anagnostopoulos  
Editor

Socratic, Platonic  
and Aristotelian Studies:  
Essays in Honor  
of Gerasimos Santas

 Springer

*Editor*

Georgios Anagnostopoulos  
Department of Philosophy  
University of California-San Diego  
La Jolla, CA 92093-0119, USA  
ganagnostopoulos@ucsd.edu

ISBN 978-94-007-1729-9

e-ISBN 978-94-007-1730-5

DOI 10.1007/978-94-007-1730-5

Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011934286

© Springer Science+Business Media B.V. 2011

No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))



Gerasimos Santas

# Contents

<b>Acknowledgements</b> . . . . .	ix
<b>Abbreviations of Plato's Works</b> . . . . .	xi
<b>Abbreviations of Aristotle's Works</b> . . . . .	xiii
<b>Contributors</b> . . . . .	xv
<b>About the Authors</b> . . . . .	xvii
<b>Introduction</b> . . . . .	xxiii
<b>1 The Diagnostic Function of Socratic Definitions</b> . . . . .	1
Michael Ferejohn	
<b>2 Definition and Elenchus</b> . . . . .	19
Nicholas White	
<b>3 Reasons and the Problem of the Socratic Elenchos</b> . . . . .	35
Alejandro Santana	
<b>4 Santas, Socrates, and Induction</b> . . . . .	53
Mark L. McPherran	
<b>5 Socrates <i>Mythologikos</i></b> . . . . .	75
Fred D. Miller	
<b>6 Is the Prudential Paradox in the <i>Meno</i>?</b> . . . . .	93
Thomas C. Brickhouse and Nicholas D. Smith	
<b>7 Gerasimos</b> . . . . .	103
Terry Penner	
<b>8 Beyond <i>De Re</i>: Toward a Dominance Theory of Desire Attribution</b> . . . . .	125
Naomi Reshotko	
<b>9 The Good and the Just in Plato's <i>Gorgias</i></b> . . . . .	145
Christopher Rowe	

<b>10</b>	<b>Socrates, Wisdom and Pedagogy</b> . . . . .	165
	George Rudebusch	
<b>11</b>	<b>The <i>Republic</i> as Philosophical Drama</b> . . . . .	185
	John P. Anton	
<b>12</b>	<b>Function, Ability and Desire in Plato's <i>Republic</i></b> . . . . .	205
	Antonis Coumoundouros and Ronald Polansky	
<b>13</b>	<b>Knowledge, Virtue, and Method in <i>Republic</i> 471c–502c</b> . . . . .	219
	Hugh H. Benson	
<b>14</b>	<b>Reasoning About Justice in Plato's <i>Republic</i></b> . . . . .	245
	Anthony W. Price	
<b>15</b>	<b>Plato on Justice</b> . . . . .	255
	David Keyt	
<b>16</b>	<b>Plato on the Ideal of Justice and Human Happiness: Return to the Cave (<i>Republic</i> 519e–521b)</b> . . . . .	271
	Yuji Kurihara	
<b>17</b>	<b>Surpassing in Dignity and Power: The Metaphysics of Goodness in Plato's <i>Republic</i></b> . . . . .	281
	Christopher Shields	
<b>18</b>	<b>Comments on Plato's Causal Explanation</b> . . . . .	297
	D.Z. Andriopoulos	
<b>19</b>	<b>Desires and Faculties in Plato and Aristotle</b> . . . . .	323
	Deborah K.W. Modrak	
<b>20</b>	<b>Is Aristotle's Function Argument Fallacious?</b> . . . . .	335
	Gavin Lawrence	
<b>21</b>	<b>Aristotle on Discovering and Desiring the Real Good</b> . . . . .	367
	Mariana Anagnostopoulos	
<b>22</b>	<b>Continuity and Incommensurability in Ancient Greek Philosophy and Mathematics</b> . . . . .	389
	Vassilis Karasmanis	
	<b>The Works of Gerasimos Santas</b> . . . . .	401
	<b>Index</b> . . . . .	405



## Acknowledgements

Editing this volume has been both a pleasure and a learning experience. It gave me the opportunity to reconnect with some colleagues I have known for years and to become acquainted with additional ones, all of whom share an admiration for Professor Gerasimos (Jerry) Santas' work on ancient Greek philosophy. Interacting with them and dealing with editorial matters in connection with their contributions proved, contrary to accepted wisdom, an enjoyable undertaking. At the same time, while working on the volume I learned much from the essays within it about the thought of Socrates, Plato, and Aristotle. All the authors make important contributions to the study of ancient philosophy, and some of them make considerable advances in the scholarly debates on the thought of these three seminal thinkers. I want to thank all the authors for accepting the invitation to participate in the project, for their cooperation throughout the editorial process and, above all, for their contributions. I am also grateful to them for their perseverance. The first steps toward publishing these essays in honor of Professor Santas started in 2006, but for reasons that none could have foreseen or even imagined at that time, it has taken years to finish the project. I appreciate their support in keeping the project alive. (Earlier versions of the essays in this volume appeared in two issues of *Philosophical Inquiry*, Vol. XXX, No. 3–4, 2008 and Vol. XXXI, No. 1–2, 2009.)

Special thanks are owed to a number of people who have been most helpful with the publication of this volume: the editor of Springer's *Philosophical Studies Series*, Professor Stephen Hetherington, for his encouragement and decision to consider the volume for the series; Professor Nicholas D. Smith for all his efforts throughout the process of finding a suitable publisher for the volume; an anonymous referee for an exhaustive review of the manuscript; Ms. Ingrid van Laarhoven and Ms. Hendrikje Tuerlings at Springer for all their assistance in the publication of the volume; Ms. Karen Jessie for guiding the production of the volume and Craig Aguille for assisting with the proofs and index. It has been a special pleasure to work with all of them.

Lastly, I would like to thank Jerry Santas for his invaluable assistance throughout the planning and preparation of the volume, but most of all for keeping the faith and helping to bring the project to a happy conclusion.

It is with much affection that I and all the contributors dedicate this volume to Jerry, in recognition and appreciation of the major contributions he has made to the study of ancient Greek philosophy, his professional ethos, and his warm collegiality over an academic career that spans half a century.

Del Mar, California  
March 25, 2011

Georgios Anagnostopoulos

# Abbreviations of Plato's Works

<i>Alc.</i>	<i>Alcibiades</i>
<i>Ap.</i>	<i>Apology</i>
<i>Chrm.</i>	<i>Charmides</i>
<i>Cleit.</i>	<i>Cleitophon</i>
<i>Cra.</i>	<i>Cratylus</i>
<i>Criti.</i>	<i>Critias</i>
<i>Epin.</i>	<i>Epinomis</i>
<i>Epist.</i>	<i>Epistles</i>
<i>Euphr.</i>	<i>Euthyphro</i>
<i>Euthd.</i>	<i>Euthydemus</i>
<i>Grg.</i>	<i>Gorgias</i>
<i>H. Ma.</i>	<i>Hippias Major</i>
<i>H. Mi.</i>	<i>Hippias Minor</i>
<i>La.</i>	<i>Laches</i>
<i>Lys.</i>	<i>Lysis</i>
<i>Menex.</i>	<i>Menexenus</i>
<i>Phd.</i>	<i>Phaedo</i>
<i>Phdr.</i>	<i>Phaedrus</i>
<i>Phil.</i>	<i>Philebus</i>
<i>Pol.</i>	<i>Politicus (Statesman)</i>
<i>Prm.</i>	<i>Parmenides</i>
<i>Prt.</i>	<i>Protagoras</i>
<i>Rep.</i>	<i>Republic</i>
<i>Smp.</i>	<i>Symposium</i>
<i>Sph.</i>	<i>Sophist</i>
<i>Theag.</i>	<i>Theages</i>
<i>Tht.</i>	<i>Theaetetus</i>
<i>Ti.</i>	<i>Timaeus</i>

## Abbreviations of Aristotle's Works

<i>An</i>	<i>On the Soul (de Anima)</i>
<i>An. Pr</i>	<i>Prior Analytics (Analytica Priora)</i>
<i>An. Post</i>	<i>Posterior Analytics (Analytica Posteriora)</i>
<i>Cael</i>	<i>On the Heavens (de Caelo)</i>
<i>Cat</i>	<i>Categories (Categoriae)</i>
<i>EE</i>	<i>Eudemian Ethics (Ethica Eudemia)</i>
<i>GC</i>	<i>On Generation and Corruption (de Generatione et Corruptione)</i>
<i>HA</i>	<i>History of Animals (Historia Animalium)</i>
<i>IA</i>	<i>Progression of Animals (de Incessu Animalium)</i>
<i>Int</i>	<i>Interpretations (de Interpretatione)</i>
<i>MA</i>	<i>Movement of Animals (de Motu Animalium)</i>
<i>Mem</i>	<i>On Memory (de Memoria et Reminiscentia)</i>
<i>Met</i>	<i>Metaphysics (Metaphysica)</i>
<i>Meteor</i>	<i>Meteorology (Meteorologica)</i>
<i>MM</i>	<i>Magna Moralia</i>
<i>Mund</i>	<i>On the Universe (de Mundo)</i>
<i>NE</i>	<i>Nicomachean Ethics (Ethica Nicomachea)</i>
<i>PA</i>	<i>Parts of Animals (de Partibus Animalium)</i>
<i>Phys</i>	<i>Physics (Physica)</i>
<i>Poet</i>	<i>Poetics (Poetica)</i>
<i>Pol</i>	<i>Politics (Politica)</i>
<i>Prob</i>	<i>Problems (Problemata)</i>
<i>Rhet</i>	<i>Rhetoric (Rhetorica)</i>
<i>SE</i>	<i>Sophistical Refutations (Sophistici Elenchi)</i>
<i>Sens</i>	<i>Sense and Sensibilia (de Sensu et Sensibilibus)</i>
<i>Top</i>	<i>Topics (Topica)</i>
<i>VV</i>	<i>On Virtues and Vices (de Vertutibus et Vitiis)</i>

# Contributors

**Mariana Anagnostopoulos** California State University, Fresno, CA, USA,  
marianaa@csufresno.edu

**D.Z. Andriopoulos** Aristotle University of Thessaloniki, Thessaloniki, Greece

**John P. Anton** University of South Florida, Tampa, FL, USA,  
hanton1@tampabay.rr.com

**Hugh H. Benson** University of Oklahoma, Norman, OK, USA, hbenson@ou.edu

**Thomas C. Brickhouse** Lynchburg College, Lynchburg, VA, USA,  
brickhouse@lynchburg.edu

**Antonis Coumoundouros** Adrian College, Adrian, MI, USA,  
acoumoundouros@adrian.edu

**Michael Ferejohn** Duke University, Durham, NC, USA, mtf@duke.edu

**Vassilis Karasmanis** National Technical University of Athens, Athens, Greece,  
vasmar@hol.gr

**David Keyt** University of Washington, Seattle, WA, USA,  
keyt@u.washington.edu

**Yuji Kurihara** Tokyo Gakugei University, Koganei, Tokyo, Japan,  
ykuri@u-gakugei.ac.jp

**Gavin Lawrence** University of California, Los Angeles, CA, USA,  
lawrence@humnet.ucla.edu

**Mark L. McPherran** Simon Fraser University, Burnaby, BC, Canada,  
mark\_mcphehrran@sfu.ca

**Fred D. Miller, Jr.** Bowling Green State University, Bowling Green, OH, USA,  
fmiller@bgsu.edu

**Deborah K.W. Modrak** University of Rochester, Rochester, NY, USA,  
deborah.modrak@rochester.edu

**Terry Penner** University of Wisconsin-Madison, Madison, WI, USA,  
tmpenner@facstaff.wisc.edu

**Ronald Polansky** Duquesne University, Pittsburgh, PA, USA, polansky@duq.edu

**Anthony W. Price** Birkbeck College, London, UK, a.price@bbk.ac.uk

**Naomi Reshotko** University of Denver, Denver, CO, USA,  
Naomi.Reshotko@du.edu

**Christopher Rowe** Durham University, Durham, UK, c.j.rowe@dur.ac.uk

**George Rudebusch** Northern Arizona University, Flagstaff, AZ, USA,  
George.Rudebusch@nau.edu

**Alejandro Santana** University of Portland, Portland, OR, USA, santana@up.edu

**Christopher Shields** University of Oxford, Oxford, UK,  
christopher.shields@lmh.ox.ac.uk

**Nicholas D. Smith** Lewis & Clark College, Portland, OR, USA,  
ndsmith@lclark.edu

**Nicholas White** University of Utah, Salt Lake City, UT, USA,  
nicholas.white@philosophy.utah.edu

## About the Authors

**Georgios Anagnostopoulos** is a Professor of philosophy at the University of California, San Diego. He has authored *Aristotle on the Goals and Exactness of Ethics* (The University of California Press, 1994), edited *A Companion to Aristotle* (Wiley-Blackwell, 2009) and *Law and Rights in the Ancient Greek Tradition* (a Supplementary Volume of *Philosophical Inquiry*, 2006), and has published a number of articles on ancient Greek philosophy and medicine.

**Mariana Anagnostopoulos** is Assistant Professor of Philosophy at California State University, Fresno. She is the author of “The Divided Soul and the Desire for Good in Plato’s *Republic*” (in *The Blackwell Guide to Plato’s Republic*, 2006), among other papers on Socratic, Platonic, and Aristotelian theories of ethics and moral psychology. Her teaching interests also include the history and application of ethics, and twentieth-century analytic philosophy.

**D.Z. Andriopoulos** is Professor of Philosophy Emeritus at the Aristotle University of Thessaloniki. He has authored *Greek Theory of Knowledge* (Hestia 1987; reprinted 1995, 2002), *Causality in Ancient Greek Philosophy* (Aristotle University Press 1995; reprinted 2002, 2003), *Sense and Perception in Greek Philosophy* (W. Green 1974), *Neo-Hellenic Aesthetic Theories* (Kampanas 1977), *History of Aesthetics Vol. I, II* (Aristotle University Press 1990; reprinted 1996, 2002); has edited *Plato: Fifty Essays, Vol. I, II, III* (Papademas 1998; reprinted 2001), *Aristotle: Forty Essays in Honor of J. P. Anton* (Aristotle University Press 1993; reprinted 1997, 2003); and has published numerous articles. He is the founder and editor from 1978 to the present of the international quarterly, *Philosophical Inquiry*, in which Professor G. Santas has published several of his seminal papers on Plato and Aristotle.

**John P. Anton** is Distinguished Professor of Greek Philosophy and Culture at the University of South Florida. Among his books are *Aristotle’s Theory of Contrariety, Categories and Experience* and *American Naturalism and Greek Philosophy*. He is also a Corresponding Member of the Academy of Athens.

**Hugh H. Benson** is Professor and Chair of the Department of Philosophy at the University of Oklahoma. He was a Samuel Roberts Noble Presidential Professor from 2000 to 2004. He is the editor of *Essays on the Philosophy of Socrates* (Oxford,

1992) and *A Companion to Plato* (Blackwell 2006), and author of *Socratic Wisdom* (Oxford, 2000) as well as various articles on the philosophy of Socrates, Plato, and Aristotle.

**Thomas C. Brickhouse** is currently the John M. Turner Chair in the Humanities and Professor of Philosophy at Lynchburg College. He has often collaborated with Nicholas D. Smith on work on the philosophy of Socrates. Their latest co-authored book, *Socrates' Moral Psychology*, will appear with Cambridge next year. For his other co-authored books, see the entry on Nicholas D. Smith.

**Antonis Coumoundouros** is Assistant Professor of Philosophy at Adrian College in Michigan, where he teaches courses in ancient philosophy and the history of philosophy. He wrote his dissertation on Plato's view of tyranny and has published articles on ancient political philosophy and ethics. He is currently working on an introduction and translation of Plato's *Critias*.

**Michael Ferejohn** is an Associate Professor of Philosophy at Duke University, specializing in ancient philosophy, metaphysics, epistemology, and philosophical logic. He is the author of *The Origins of Aristotelian Science* (Yale, 1991), and has published numerous articles on epistemology and metaphysics in Plato and Aristotle.

**Vassilis Karasmanis** (D.Phil. Oxford) is a Professor of Philosophy at the National Technical University of Athens. He is the author of numerous articles on ancient philosophy and ancient mathematics. Recently, he co-edited (with Lindsay Judson) *Remembering Socrates* (Oxford, 2006).

**David Keyt** is a Professor of philosophy at the University of Washington. In the 1950s he and Jerry Santas were graduate students together in the Sage School of Philosophy at Cornell and then toward the end of the century he and Jerry were colleagues for a brief period when he was a visiting professor at the University of California at Irvine. During his entire academic career he has been an avid student of Jerry's work on Plato and Aristotle. He is the author of *Aristotle: Politics V and VI* (Clarendon Press, 1999) and co-editor with Fred D. Miller, Jr. of *A Companion to Aristotle's Politics* (Blackwell, 1991) and *Freedom, Reason, and the Polis: Essays in Ancient Greek Political Philosophy* (Cambridge, 2007).

**Yuji Kurihara** received his Ph.D. from the University of California, Irvine, in Ancient Philosophy, and is Associate Professor of Philosophy and Ethics at Tokyo Gakugei University. He has published several articles on Plato's epistemology and ethics in Japanese and English.

**Gavin Lawrence** is a Professor of Philosophy at the University of California, Los Angeles. He is the author of a number of articles on Aristotle's ethics, and on modern ethics and practical philosophy.

**Mark L. McPherran** is a Professor and Chair of Philosophy at Simon Fraser University. He is the author of *The Religion of Socrates* (Pennsylvania State University Press, 1996), the editor of *Wisdom, Ignorance, and Virtue: New Essays in*



*Socratic Studies* (1997) and *Recognition, Remembrance, and Reality: New Essays on Plato's Epistemology and Metaphysics* (1999), and a variety of articles on Socrates, Plato, and ancient skepticism. Recent work includes: "Socratic Religion," in *The Cambridge Companion to Socrates* (forthcoming); "Medicine, Magic, and Religion in Plato's *Symposium*," in *Plato's Symposium: Issues in Interpretation and Reception* (Harvard, 2006); "Platonic Religion," in *A Companion to Plato* (Blackwell, 2006); and "The Piety and Gods of Plato's *Republic*," in *The Blackwell Guide to Plato's Republic* (Blackwell, 2006). He is the Director of the Annual Arizona Colloquium in Ancient Philosophy.

**Fred D. Miller, Jr.** is Professor of Philosophy and Executive Director of the Social Philosophy and Policy Center at Bowling Green State University. He is the author of *Nature, Justice, and Rights in Aristotle's Politics* (Oxford University Press, 1995) and co-editor of numerous collections, including *A Companion to Aristotle's Politics* (Blackwell, 1991), *Freedom, Reason, and the Polis: Essays in Ancient Greek Political Philosophy* (Cambridge University Press, 2007), and *A History of the Philosophy of Law from the Ancient Greeks to the Scholastics*, Volume 6 of *A Treatise of Legal Philosophy and General Jurisprudence* (Springer, 2007). He served as President of the Society for Ancient Greek Philosophy 1988–2004.

**Deborah K.W. Modrak** is the author of numerous papers on Ancient Greek cognitive theories and philosophy of mind and language, and has also published two books, *Aristotle on the Power of Perception* (Chicago, 1987) and *Aristotle's Theory of Language and Meaning* (Cambridge, 2001). She is a Professor of Philosophy at the University of Rochester.

**Ronald Polansky** is Professor of Philosophy at Duquesne University. He edits the journal *Ancient Philosophy*. He has published *Aristotle's De anima* (Cambridge University Press) and *Philosophy and Knowledge: A Commentary on Plato's Theaetetus* (Bucknell University Press). Besides work on ancient philosophy, he has published on bioethics and early modern philosophy.

**Terry Penner** is Professor of Philosophy Emeritus at the University of Wisconsin-Madison. In 2004, he gave the C.J. de Vogel lecture at the 7th triennial conference of the International Plato Society, and in 2005, he was the fourth biennial A. G. Leventis Research Professor of Greek in the University of Edinburgh. His publications include *The Ascent from Nominalism* (1987), *Plato's Lysis* (Cambridge 2005, with Christopher Rowe), and several essays in *Pursuing the Good: Ethics and Metaphysics in Plato's Republic* (Edinburgh 2007). His articles on Socrates include "The Death of the so-called 'Socratic Elenchus'" (International Plato Society VII, 2007), "Socrates and the Impossibility of Belief-Relative Sciences" (Boston Area Colloquium in Ancient Philosophy III, 1988), "Platonic Justice and what we mean by 'justice'" (2005, at [www.nd.edu/~plato/plato5issue/Penner.pdf](http://www.nd.edu/~plato/plato5issue/Penner.pdf)), and articles in various Cambridge and Blackwell's *Companions*.

**Anthony W. Price** is Reader in Philosophy at Birkbeck College, London. He has published three books: *Love and Friendship in Plato and Aristotle* (Clarendon Press 1989/97), *Mental Conflict* (Routledge, 1995), and *Contextuality in Practical Reason* (Clarendon Press, 2008). He enjoyed a Leverhulme Research Fellowship in 2006-8, out of which should emerge a monograph on practical reason in Plato and Aristotle.

**Naomi Reshotko** is Professor and Chair of Philosophy at the University of Denver. She is the author of *Socratic Virtue: Making the Best of the Neither-Good-Nor-Bad* (Cambridge, 2006) and has made many contributions to the literature on Socratic Ethics and Platonic Metaphysics and Epistemology.

**Christopher Rowe** is a Professor in the Department of Classics and Ancient History, Durham University, UK. His publications include *The Cambridge History of Greek and Roman Political Thought* (edited, with Malcolm Schofield, Cambridge University Press, 2000), *Aristotle, Nicomachean Ethics* (philosophical commentary by Sarah Broadie, translation by Christopher Rowe, Oxford University Press, 2002), *Plato's Lysis* (with Terry Penner, Cambridge University Press, 2005), and *Plato and the Art of Philosophical Writing* (Cambridge University Press, 2007). He is currently working on a new edition of the Penguin volume *The Last Days of Socrates*.

**George Rudebusch** is a Professor of philosophy at Northern Arizona University. He is the author of *Socrates, Pleasure, and Value* (Oxford, 1999) and *Socrates* (Wiley-Blackwell, 2009).

**Alejandro Santana** is an Associate Professor of Philosophy at the University of Portland. He is the author of "Constructivism and the Problem of the Socratic Elenchos," which appeared in the Fall 2007 issue of *Ancient Philosophy*. Along with his research in Socratic methodology, Dr. Santana also conducts research in Aztec thought. He recently published "Did the Aztecs Do Philosophy?" which appeared in the Fall 2008 issue of the *APA Newsletter on Hispanic/Latino Issues in Philosophy*. His interests also include Socio-political philosophy, Ethics, and Logic.

**Christopher Shields** is a Fellow of Lady Margaret Hall and Professor of Classical Philosophy in the University of Oxford. He is the author of *Aristotle* (Routledge), *Order in Multiplicity: Homonymy in the Philosophy of Aristotle* (Oxford), *Aristotle, De Anima: Translation and Commentary* (Oxford), *Classical Philosophy: A Contemporary Introduction* (Routledge), and (with Robert Pasnau) *The Philosophy of Thomas Aquinas* (Westview). He is also editor of *The Blackwell Guide to Ancient Philosophy* (Blackwell) and *The Oxford Handbook on Aristotle* (Oxford).

**Nicholas D. Smith** is the James F. Miller Professor of Humanities in the Philosophy Department at Lewis & Clark College in Portland, Oregon. His other collaborations with Thomas C. Brickhouse include: *Socrates on Trial* (Oxford and Princeton, 1989), *Plato's Socrates* (Oxford, 1994), *The Philosophy of Socrates* (Westview, 2000), *The Trial and Execution of Socrates: Sources and Controversies* (co-edited; Oxford, 2002), and the *Routledge Philosophy Guidebook to Plato and the Trial*

*of Socrates* (Routledge, 2004). Their latest co-authored book, *Socrates' Moral Psychology* will appear with Cambridge next year.

**Nicholas White** is Presidential Professor at the University of Utah, and Professor Emeritus at the University of Michigan, Ann Arbor, and the University of California at Irvine. He is the author of the books *Plato on Knowledge and Reality*, *A Companion to Plato's Republic*, *Individual and Conflict in Greek Ethics*, and *A Brief History of Happiness*.

# Introduction

Georgios Anagnostopoulos

The contributions of Gerasimos Santas to the study of ancient Greek philosophy are well known to scholars working on Socrates, Plato, and Aristotle. Having started in the 1960s with a number of seminal papers on the Socratic Paradoxes and Socrates' view on the possibility of the weakness of will, he authored his award-winning book on Socrates in the Arguments of the Philosophers Series (Routledge 1979). These early writings had a major impact with respect to the interpretations of the Platonic texts they advanced. But that was not their only influence in Socratic scholarship. They also shifted the focus from the historical debates about Socrates and the perennial "Socratic Problem" to the philosophical views of Socrates as they are presented in the Early or Socratic Dialogues of Plato. Vlastos acknowledged this effect of Santas' approach to the study of Socrates, claiming that it liberated Socratic scholarship from what he called the "bugbear of Platonic studies, the so-called 'Socratic Problem'" and confessed to being strongly attracted to it (Vlastos, p. 45).

As Socratic scholars know, many of the views attributed to Socrates seem not merely counterintuitive but almost paradoxical. His alleged denial of the possibility of weakness of the will (*akrasia*) is perhaps the best example; it is considered by many counterintuitive, if not downright false, while some of his positions on prudential and moral motivation have been labeled "Prudential Paradoxes" and "Moral Paradoxes" respectively. Santas' early work has provided some of the most clear, insightful, and influential analyses of Socrates' views on these matters and has shed much light on the Socratic puzzles. Speaking on the importance of these papers, Vlastos writes that "For me the turn-around began with Santas' paper on "The Socratic Paradox" in 1964, which first shed light on one of the thickets of Socratic strangeness, making good sense of a thesis which Socrates had long seemed to me to be asserting in mulish defiance of common sense" (Vlastos, p. 19). Among other things, Santas' application of the distinction between the intended and actual object of desire—a distinction that derives from Frege's well known account of reference in opaque or intentional contexts—provided a way of making sense of the Socratic puzzles about desire and motivation in the *Meno* and *Gorgias*. Many scholars became convinced of the plausibility of Santas' interpretations of the prudential and moral paradoxes, including Vlastos who concluded that Santas had used the distinction to resolve the Socratic paradoxes (Vlastos, p. 151). But as is

common in scholarly debates, not all were as convinced as Vlastos, and Santas' interpretations generated a lively debate about the Socratic views on desire and motivation. T. Penner, who rejects the Fregean account of reference in intentional contexts, has offered the most systematic and insightful alternative to Santas' views (see his contribution to this volume and his papers listed in the Bibliography at the end of his contribution; see also, the contributions of T. Brickhouse and N. Smith, N. Reshotko, and M. Anagnostopoulos in this volume). Similarly, Santas' analysis of Socrates' account of the power of knowledge in prudential and ethical conflicts in terms of the strength of desires offered an original interpretation of the counterintuitive Socratic position on *akrasia*, making this position much more plausible than it had seemed.

Many of these early papers, and additional ones about Socrates' conceptions of virtue in general as well as of some of the specific virtues, became a part of his book on Socrates. The book is probably the first systematic account of the philosophical views of Socrates that simultaneously offers a rigorous formulation and assessment of the arguments Socrates gives in support of these views. Santas was correct in seeing even dialogues such as the *Apology* and *Crito* as expounding and defending in typical Socratic arguments Socrates' philosophical positions—e.g., that it is never right to harm anyone or to violate a just agreement. His reconstructions of the Socratic views and arguments in these dialogues are models of clarity and rigor, with respect to both textual interpretation and philosophical analysis. Santas' systematic discussion in Parts I and III of his book of Socrates' normative ethical principles, his views on the nature of virtue, and his allegedly counterintuitive claims about motivation, *akrasia*, and the object of desire showed that the various Socratic claims could be seen as parts of a comprehensive theory of ethics, motivation, and action that aspired to being coherent. But the contributions of Santas' book go beyond the ones just discussed.

One third (Part II) of Santas' book on Socrates is devoted to an examination of Socratic method, primarily Socrates' mode of argument and concern with definition, which together comprise what is often thought to be the very heart of Socratic practice. Santas methodically isolates and systematically catalogues all the definitions that are to be found in Plato's Socratic Dialogues, whether proposed by Socrates' interlocutors or Socrates himself. He then examines their syntactic, pragmatic, and semantic features and identifies their actual or intended uses within the Socratic practice—e.g., diagnostic and epistemic uses. One effect of this detailed study of the Platonic texts with respect to Socratic definitions and their role in the Socratic practice was the putting to rest of much speculation about the supposed real aims of Socratic inquiry and of the doubts frequently voiced in the scholarly debates as to whether Socrates had any genuine interest in definitions at all. In addition, Santas' illuminating study of Socrates' elenctic reasoning, of its use both of deduction and *epagôgê*, went a long way in renewing interest in the study of the Socratic *elenchos* and the central role it plays in Socratic practice (see the contributions of J. Ferejohn, M. McPherran, and A. Santana in this volume).

Santas' interest in and understanding of the Socratic/Platonic accounts of desire and, in general, Socratic/Platonic moral psychology led to his much admired study

of Plato's and Freud's theories of love (Blackwell 1988). After setting out the aim of and constraints on a theory of love, Santas begins his account of Plato's views on love by offering one of the most illuminating interpretations of the central theme of the *Symposium*—the nature of *eros*—that goes a long way toward resolving a number of apparent and real inconsistencies in the various accounts of love presented in that dialogue. In his analysis of the underlying structure of the motivation of love assumed by the speakers in the *Symposium*, Santas points out that two potentially conflicting models of love are put forth, the “deficiency model” and “egoistic model.” According to the first model, which is put forth by most of the *Symposium* speakers, love stems from lack, need, or deficiency in the lover. According to the second, which is mainly defended by Agathon (although Socrates used to think like Agathon on this point) and which Santas sees as narcissistic in character, those who are themselves good and beautiful are attracted to the same qualities in others. Santas points out that these two models seem inconsistent: If we love only what we lack, as the “deficiency model” asserts, how can we love ourselves? On the other hand, if we love what we possess, as the “egoistic model” asserts, how can we be said to lack it? Santas finds the resolution of the conflict in the ingenious hypothesis Diotima proposes in her *Symposium* speech—those who love are not completely good, wise, and beautiful but only moderately so. She argues that, if they were completely ignorant and bad, they would be unaware of what they lacked and would have no desire for it. Since they possess these characteristics to a moderate degree, they appreciate them and desire to enjoy them more. Santas argues that Diotima's way of resolving the conflict—by accepting a deficiency of some quality in the lover, but not total lack of it—has an important implication about the possibility of the gods loving anything. If the gods are perfectly virtuous and lack nothing, they cannot love anything. According to Santas, this implication figures prominently in Socrates' argument leading to the conclusion that *Eros* is not a god.

Santas' discussion of a different puzzle and its resolution in the *Symposium* is equally illuminating. Most often *eros* is understood as being a desire for beauty. But at times it is asserted, e.g., by Diotima, that it is a desire for the good, to possess it forever, or a desire for immortality. Relying on ideas he develops in his seminal paper on Plato's conception of the Good in the *Republic* (see below), Santas distinguishes between “generic *eros*,” which is a desire for the good and immortality, and “proper *eros*,” which is a desire for beauty. Since *eros* proper is according to Plato a species of generic *eros*, the object of the *eros* proper, beauty, can be thought of as being included in the object of generic *eros*, the good.

Santas sees yet another problem in the *Symposium*, namely, that love in the highest steps of the ladder of love—e.g., love for Beauty, souls, laws, etc.—is viewed as asexual, which raises the question of whether love for such objects is *eros* at all. He argues that the theories of the soul and motivation articulated in the *Phaedrus* are an advance over what we encounter in the *Symposium* and imply that all human love involves sexual passion. In the *Phaedrus*, the human psyche is seen as always being pulled by two horses, the one, representing reason, pulling upwards to associate with the Forms and the other, representing appetite, pulling in the opposite direction in pursuit of pleasure. This account of the human psyche and desire implies that love

always has a sexual aspect. In his discussion of the *philia*, the kind of love that is the subject of *Lysis*, Santas sees in the Socratic claim that “what is neither good nor bad is friend to the good” the “deficiency model” of *eros* as amended by Diotima at work once again. But the application of the “deficiency model,” even as amended by someone as wise as Diotima, has its difficulties as an account of *philia*. Santas points out that if *philia* is mutual and its nature is what the deficiency model makes it out to be, then *philia* is not possible: the good person has no need to be a friend of or love anyone who is either good or bad. Santas argues that the deficiency model is abandoned in the *Republic*, when Plato presents his views about friendship among rulers and guardians as a kind of sharing among good people of what they enjoy.

Santas’ discussion of Freud’s views on love is as insightful as that of the views of Plato. He points out that at least two distinct accounts of love can be identified in Freud’s writings—the one Freud held until 1915 and called “anaclitic” and a later one that followed his paper on narcissism in that same year and might be called the “narcissistic” account. Freud’s early view explains adult love in terms of the childhood attachment that presumably a child has for the parent of the opposite sex; in adult love a new person replaces the parent and becomes the object of sexual desire and affection. In Freud’s later view, formulated primarily with reference to women, love stems from love for oneself. A woman loves her children as an extension of herself and loves others by a kind of transference. Santas’ discussion sheds much light on Freud’s different and seemingly inconsistent accounts of love. He also makes many insightful observations about Freud’s account of sublimation, according to which all love, even what seems not to aim at sexual pleasure, has its origin in the libido and aims at gratification. Santas points out the clear differences between Platonic *eros* and Freudian libido, and is critical of Freud’s claim that they are the same. His comparison of the theories of the two thinkers and the assessment of each one of them in terms of what he had earlier identified as the conditions a theory of love must meet result in many original and valuable observations about the strengths and weaknesses of the Platonic and Freudian accounts of love.

In his more recent book, *Goodness and Justice* (Blackwell 2001), Santas focuses on the two most central elements of ancient Greek theories—goodness and justice—especially the former, which he takes to be the most fundamental concept in Plato’s and Aristotle’s ethics and the one in terms of which they “*build up* their theories of virtue, especially justice, and of happiness” (p. 16). One of the many strengths of the book is Santas’ insistence on examining the views of the ancients as a part of the long history of ethics, especially its Modern part, and seeing the views of Plato and Aristotle against those of Hume, Kant, Sidgwick, Moore, and especially Rawls. Many valuable insights result from these comparisons, especially the length to which the ancients went in order to refute subjective theories of the good and defend alternatives that passed some kind of objectivity test. Other strengths of the book are the emphasis Santas places on as well as his success in articulating the form and structure of the ethical theories that are the focus of his study. This makes it possible for him to examine at an unusually deep level the perplexing questions as to whether the ethical theories of Plato and Aristotle belong, by their



structure, to the teleological type or whether they represent the first examples of virtue-ethics theories. To consider two examples, Santas argues that a basic concept in Plato's ethical theory is that of functioning well, and that this concept provides him with a conception of the good that is prior to and independent of virtue. But without a maximizing principle, this does not make Plato's theory strictly teleological. Santas argues that Plato has such a maximization principle for justice in the *polis* (social or political justice) but not for justice in the soul (individual justice). Thus, with the respect to individual justice, Plato's view seems to be a type of virtue ethics. Similarly, in discussing Aristotle's views on the same issues, Santas argues that the two lines of thought about goodness—i.e., perfectionist and teleological—pose problems for clearly determining whether Aristotle's ethical theory is a type of virtue ethics, as is often claimed. Santas argues that justice and just action are ultimately determined with reference to the good of the community, so that justice is given a teleological justification. However, according to Santas, the picture is far from clear with respect to other virtues and Aristotle's views about them may be more in tune with a virtue ethics.

Santas begins his examination of the ethical theories of Plato and Aristotle with an account of the Socratic position on the ranking of goods, offers a detailed analysis of the position Socrates assigns to wisdom and virtue, and uncovers a circularity, later seen by Plato in the *Republic*, in the Socratic position that nothing apart from wisdom is unconditionally good. If wisdom is beneficial on account of making knowledge of the good possible, it cannot be the good itself. Santas argues that the circularity problem is successfully dealt with by Plato's introduction of the functional-perfectionist account of the good. This conception of the good and the conception of virtue implied by it are actually introduced by Socrates at the end of *Republic* I, a segment of the *Republic* that has all the features of a Socratic, aporetic dialogue, and which many consider not to have any strong connections to the rest of Plato's book but to be a separate work that was been bundled together with the *Republic*. Santas rejects this widely held view about the place of Book I within the *Republic*, and gives strong philosophical arguments in support of its being an integral part of Plato's ethical/political treatise.

In any case, the functional/perfectionist account of the good results in a definition of good that is independent of right, with the latter being defined by the former and maximization, thus resulting in a teleological theory. Santas carefully lays out the two conceptions of function—optimal and exclusive—and shows that the good of a functional thing, understood in terms of performing its function well, can avoid circularity by excluding any reference to moral features and can steer clear of subjectivism, since Socrates and Plato conceive of function without any essential references to human desires or interests. The functional/perfectionist conception of the good implies a conception of the virtues as the conditions necessary for a functional thing to perform its function well. Santas argues that in Books II–IV of the *Republic* Plato develops his account of social justice as a condition in which each and every citizen does not stray from the performance of the function for which he/she is suited by nature. Plato, according to Santas, follows the same strategy in his quest for an account of individual or psychic justice, which requires the kind



of strong isomorphism between city and soul that Plato aims to establish. In his discussion of the parts of the soul, Santas raises serious doubts about a view that has become prevalent in recent years. This view attributes desires to the rational and spirited parts of the soul, on the one hand, and some rational capacity to the appetitive part, on the other. He gives strong reasons for denying any cognitive activity to appetite, and offers some insightful comparisons of Plato's and Hume's views on reason and the passions.

Santas argues that Plato introduces another conception of good in the *Republic*, one that apparently different from, and independent of, the conception of good just discussed. This second conception is based on the Platonic theory of Forms, and especially Plato's notoriously cryptic account of the Form of the Good, and he labels it "metaphysical." In his seminal paper on the Form of the Good I mentioned earlier, Santas has offered an account of the Platonic Form of the Good. Santas' account has been recognized as the most insightful interpretation of that part of the *Republic*, which has seemed to many to defy interpretation and understanding (see the contribution of C. Shields in this volume). Santas distinguished between two kinds of properties a Form has—ideal ones, which a Form has by being a Form (i.e., perfection, eternity, indivisibility, knowability, immutability), and proper ones, which a Form has by being the specific Form it is. The Form of the Good is a kind of Form of Forms, or meta-Form, in which all the Forms participate in so far as they have the ideal properties that Plato attributes to them. Santas then offers a way of bringing the two apparently distinct conceptions of good in the *Republic*, the functional/perfectionist and the metaphysical, together. He argues that the function of some particular thing,  $x$ , and how well  $x$  performs this function, depend on the Form in which  $x$  participates and on the degree to which  $x$  resembles it (its degree of perfection). Thus, according to Santas, the two theories of good can be seen as one theory, "the first great perfectionist theory of goodness." (p. 191)

Santas begins his discussion of Aristotle by examining his criticisms in the early parts of the *NE* and elsewhere of Plato's views on the good. Santas' reconstruction of both the Platonic position and Aristotle's arguments against it reveal that the latter do not pose any serious threat to the Platonic position. He then proceeds to examine the various conceptions of good Aristotle introduces and his move from an *orec-tic* to a functional/perfectionist account of good, which nonetheless gives a much more prominent place to pleasure in his account of the good than Plato allows in his own. The discussion on Aristotle offers a lucid account of the structure of Aristotle's ethical theory, which enables Santas to address the perplexing questions about the nature of Aristotelian ethical theory to which I alluded earlier—e.g., whether it is a teleological or a virtue-ethics type of theory. But many other topics that are of central importance in Aristotle's ethical theory are insightfully treated by Santas, in particular the functional account of good in Chapter 7 and justice in Chapter 8. With respect to the former, Santas does a wonderful job of unpacking Aristotle's highly compressed defense of the contemplative life as the best and happiest life in terms of the intellect's role in determining what a human being is and its surpassing everything else in power and worth, despite its smallness in bulk (*NE X.7*). Many have criticized Aristotle's views on the contemplative life and his arguments

in support of it that are based on the nature of the intellect and its role in determining the essential nature of being human, but Santas shows that the Aristotelian position cannot be easily dismissed.

## Bibliography

- Santas, G. 1964. "The Socratic Paradoxes." *The Philosophical Review* 73.
- Santas, G. 1966. "Plato's *Protagoras* and Explanations of Weakness." *The Philosophical Review* 75.
- Santas, G. 1969a. "Socrates at Work on Virtue and Knowledge in Plato's *Laches*." *Review of Metaphysics* 22.
- Santas, G. 1969b. "Aristotle on the Explanation of Action, Practical Inference, and *Akrasia*." *Phronesis* 14.
- Santas, G. 1973. "Socrates at Work on Virtue and Knowledge in Plato's *Charmides*," In *Exegesis and Argument*, edited by E.N. Lee, et al., Supplementary Volume of *Phronesis*. Assen: Van Gorcum.
- Santas, G. 1979. *Socrates*. London: Routledge.
- Santas, G. 1980. "The Form of the Good in Plato's *Republic*." *Philosophical Inquiry* 2.
- Santas, G. 1981. "Plato on Love, Beauty and the Good," In *The Greeks and the Good Life*, edited by D. Depew. Fullerton: A Publication of California State University, Fullerton.
- Santas, G. 1985. "Two Theories of Good in Plato's *Republic*." *Archiv für Geschichte der Philosophie* 57.
- Santas, G. 1988. *Plato and Freud: Two Theories of Love*. Oxford: Blackwell.
- Santas, G. 2001. *Goodness and Justice: Plato, Aristotle and the Moderns*. Oxford: Blackwell.
- Santas, G. (ed.) 2008. *The Blackwell Guide to Plato's Republic*. Oxford: Blackwell.
- Vlastos, G. 1991. *Socrates, Ironist and Moral Philosopher*. Ithaca: Cornell University Press.

# Chapter 1

## The Diagnostic Function of Socratic Definitions

Michael Ferejohn

The ethical doctrines of Plato's earlier dialogues<sup>1</sup> are often quite aptly described as "intellectualist," but it is not always recognized that this label can convey two distinct, though equally correct, ascriptions. On the one hand, it could be taken to point to the distinctive "Socratic" *method* on display in these works, which can plausibly be seen as an adaptation of Eleatic rationalistic methodology for the purposes of Socrates' own ethical investigations. As such, the intellectualist characterization records Socrates' unwavering affirmation of the power of the human intellect to deliver ethical truths, including, most importantly, truths about the nature of human virtue. On the other hand, the very same characterization might also be taken to refer to the constellation of distinctively "intellectualist" ethical *doctrines* that inhabit these dialogues, all revolving around the pivotal idea that virtue is in the end an *intellectual* achievement, and that the appropriate development and employment of the rational faculty by itself is therefore enough to ensure ethically correct behavior. In the broadest terms, my principal aim in what follows will be to argue that even though each of these understandings of Socratic intellectualism appears perfectly reasonable and innocent when considered separately, together they form what is at best an odd disjointedness, and at worst a deep incoherency, running through Socrates' investigations into the nature of virtue in Plato's early dialogues.

### 1.1 The Diagnostic Function: Ethical Theory and Practice in the Socratic Dialogues

A good place to start is with Socrates' avowed motivation for undertaking these ethical investigations in the first place. Nowadays a purely theoretical interest in developing a correct understanding of key ethical concepts goes largely unquestioned, and whether such understanding might ever pay practical dividends by improving the ethical quality of human behavior, including that of the theorists

---

M. Ferejohn (✉)  
Duke University, Durham, NC, USA  
e-mail: mtf@duke.edu

themselves, is generally regarded as a separate issue. On this point, the prevailing contemporary attitude could not contrast more sharply with the portrait of Socrates we find presented in the early dialogues. For perhaps the most striking feature of that portrait is the extent to which Socrates is, as Gregory Vlastos once described him, a “single-minded moralist,”<sup>2</sup> which is to say that his attention is always fixed firmly and unswervingly only upon issues connected in the most direct manner to the paramount ethical question of how one should best live one’s life. As a result, he typically exhibits indifference if not outright disdain towards theoretical pursuits that do not bear on that all-important issue. This attitude is evidenced dramatically in the explanation he gives in an “autobiographical” passage at *Phaedo* 97b–99c for abandoning the study of natural philosophy in his early manhood after a brief youthful flirtation. He reports that he became disenchanted with the doctrines of Anaxagoras once he realized that they had no bearing on the issue of what is “best for each” (*to hekastô(i) beltiston*) and “good for all in common” (*koinon pasi agathon*) (*Phaedo* 98b).<sup>3</sup>

Predictably, then, virtually all the Socratic conversations represented in the early dialogues are focused exclusively on ethical topics. More specifically, the majority of these ethical investigations fall within Socrates’ signature project of seeking definitions of the five cardinal virtues—courage, justice, temperance, piety, and wisdom—the possession of which were presumed during the classical period to constitute the noblest and most admirable form of human life. However, in keeping with his “single-minded moralism,” we should expect that Socrates will not justify the time and effort he devotes to these conceptual inquiries unless he is convinced that the knowledge he is pursuing will have positive practical value. And just as expected, Socrates makes it clear early on in the *Euthyphro* that he does in fact believe precisely that. The passage in question occurs shortly after the dialogue’s namesake meets Socrates outside the courtroom and reveals that his business there is to prosecute his own father on a charge of murder based on a set of facts that it would be charitable to describe as “highly questionable” (*Euthyphro* 4a–d). Socrates then expresses both surprise and incredulity at this revelation with a pointed question:

But in the name of Zeus, Euthyphro, do you think that your knowledge of the divine, and also of the pious and the impious, is so precise (*akribos*), that these facts being so, as you say they are, you do not fear that in prosecuting your father you yourself might be doing an impious thing? (*Euthyphro* 4e)

When Euthyphro brashly replies that he has no such fear because he does indeed possess such “precise” knowledge, Socrates (who is himself under an indictment of impiety) in characteristic fashion beseeches him to teach this knowledge to Socrates himself forthwith. What is most important for present purposes, however, is Socrates’ subsequent declaration that if Euthyphro can impart to him this “precise” knowledge of the nature of piety (and impiety), he will then be able use it to *diagnose* various actions as either pious or impious:

Now teach me what [piety] is, so that I may keep my gaze fixed upon it, and by using it as a model, if anything that you or anyone else does is of this sort, I may say that it is pious, and if not, that it is not pious. (*Euthyphro* 6e)

Our principal concern here will be to consider how well this proclaimed diagnostic function of ethical definitions fits within the broader scheme of Socrates' intellectualism. But first some Prodician disambiguation will be in order. Generally speaking, an assertion that the possession of X will enable someone to accomplish some end Y might be construed as making at least one of three different sorts of claim. Perhaps the most natural understanding is that X is claimed to be *sufficient* for S's ability to do Y. Now, taken in the strictest possible sense, this would mean that X *all by itself* will allow S to do Y. But of course, there are no situations in which this obtains. If, for example, I were to issue the following wishful conditional,

"If I had twenty million dollars, I could travel to the space station,"

I would quite naturally be understood to be making some sort of sufficiency claim, but also to be assuming as fixed such factual conditions as that I will be able to pass the pre-flight physical exam, and that the Russian civilian space travel program will remain in operation. In the present context, it is highly unlikely that Socrates would hold that knowledge of what piety is will *all by itself* allow him to diagnose pious and impious actions, for the obvious reason that such diagnoses will undoubtedly also require specific factual information about the individual cases to which this knowledge is to be applied.<sup>4</sup> It is therefore more reasonable to understand the relation in question as something like "sufficient *in the circumstances*." Generalizing from the case of piety to the whole set of virtues, I shall therefore henceforth understand what I call the *sufficiency thesis* as entailing only that anyone who knows what a given virtue is, and also possesses an adequate factual description of any action, will be able to determine whether that action instantiates that virtue or not.<sup>5</sup>

Whether or not Socrates believes that the knowledge he seeks from Euthyphro is sufficient (in this qualified manner) for effective ethical diagnosis of action, he may or may not subscribe to an essentially negative *necessity thesis* to the effect that anyone who does not know what a given virtue is would *not* be able to determine which actions instantiate that virtue. This controversial thesis, which has been featured prominently as the "priority of definition" principle in recent discussions of the so-called "Socratic Fallacy,"<sup>6</sup> is not entailed, nor even suggested, by the language of *Euthyphro* 6e itself. On the other hand, it does seem to be endorsed in the earlier passage quoted above, where Socrates' expresses the incredulity that leads him to press his interlocutor for instruction on the nature of piety.

But in the name of Zeus, Euthyphro, do you think that your knowledge of the divine, and also of the pious and the impious, is so precise, that these facts being so, as you say they are, you do not fear that in prosecuting your father you yourself might be doing an impious thing? (*Euthyphro* 4e)

Admittedly, it is a tricky business to infer positive beliefs from the questions someone poses as opposed to statements they make, but in the present case it seems quite possible that the source of Socrates' amazement here is a firm conviction on his part that it would not be possible for Euthyphro to know that his prosecution—or any other action—exemplified piety *unless* he possessed the sort of "precise" knowledge in question. And if this is reasonably understood as a flowing from a general

epistemic limitation, and not some special characteristics of Euthyphro himself, it then appears that Socrates does indeed think that knowing the nature of the virtues is necessary for diagnosing virtuous actions.

The final way of understanding the diagnostic function of Socratic definitions we shall have to consider is the weakest of all. This commits him to nothing beyond what I shall call the “utility thesis,” that knowing the nature of a virtue can be useful on some occasions in diagnosing its instances. On this construal it is left entirely open (1) whether, on those occasions, such knowledge is sufficient (in the manner specified above) for successful diagnosis, and also (2) whether there might be other occasions in which the diagnosis can be accomplished in the absence of this sort of knowledge.

Having paid Prodicus his proper tribute, I now propose to explore what I believe is a deep tension between this proposed diagnostic use of definitions on one hand and Socrates’ commitment to certain “intellectualist” ethical doctrines in the early dialogues on the other. Specifically, I shall argue that even though Socrates’ proposed diagnostic use of the definitions he seeks seems reasonable enough when considered on its own, it becomes highly problematic when we take into account his distinctive “intellectualist” preconceptions about the fundamental nature of the virtues and how they should be defined.

However, even before any particular assumptions about the nature of the virtues come into play, there is some reason to suspect on general grounds that the sufficiency thesis as formulated above is implausibly strong.<sup>7</sup> If Socrates is not being ironic or disingenuous when he issues his requests for definitions, then he must believe it is at least theoretically possible for him to encounter someone who could meet his challenge by producing a correct definition and successfully defending it against his attempted refutations. Let us then suppose for the sake of argument that this actually occurs and that Socrates could then be said, in the language of *Euthyphro* 6e, to have been “taught” the definition of the virtue in question. The question now is whether it’s really plausible to think that on the basis of this alone Socrates would be in a position to diagnose every possible act as virtuous or non-virtuous. To put it another way, leaving aside the epistemic status of his imaginary interlocutor, is it not paradoxical that Socrates could be transformed into a consummate moral authority in the course of just one single elenctic encounter? On the other hand, if what Socrates is proposing at *Euthyphro* 6e and like passages is that he be given a full course of ethical study under his interlocutor’s tutelage—a course that could conceivably take years—it would be hard to defend him against the charge of misdirection when he gives his reason for seeking definitions.<sup>8</sup>

I should make clear that at this point I am raising this issue as an initial general worry about the diagnostic function of definitions, not suggesting that it yet constitutes a decisive objection to the Socratic ethical program in the early dialogues. The issue of whether definitions of the virtues can fulfill the diagnostic function described at *Euthyphro* 6e cannot be resolved in the abstract, but instead must be addressed by first determining what sort of definitions are at issue, and then asking how exactly they could be employed in diagnosing virtuous actions. In the following two sections I will follow this procedure first for definitions of the virtues in terms

of overt behavior, which I will argue Socrates systematically rejects, and then for a type of psychological definition that I will argue he advocates in place of behavioral ones.

## 1.2 Socrates' Rejection of Behavioral Definitions

In almost every case, the earliest definitions proposed by Socrates' interlocutors identify the virtue in question with some sort of *behavioral pattern*, or perhaps a disposition to exhibit such a pattern. For example, the initial definition of courage in the *Laches* is "staying at one's post and facing the enemy" (*Laches* 190e), that of temperance in the *Charmides* "a sort of quietness" (159c), and that of justice in *Republic* I "not cheating and always paying one's debts" (331b).<sup>9</sup> In each of these contexts Socrates rejects the proposed definition by pointing to cases where the behavior in question is present but the virtue is evidently lacking, or vice versa, which effectively serves to undermine co-extensiveness between definiendum and proposed definiens. Specifically, in the *Laches* (191a–b) Socrates points to cases of courageous persons (the Scythian cavalymen) who fight by skirmishing rather than in formation, in the *Charmides* (159c–160c) he argues that in many endeavors quickness is better (and so presumably more virtuous) than quietness, and in *Republic* I (331c) he raises the case of someone borrowing a spear from an acquaintance who then demands repayment after having gone mad in the interim.

It might be supposed that the correct lesson to draw from these encounters is that the actual behavioral definitions proposed in the texts are merely too crude, and that what is needed is simply some fine-tuning to align the definiendum and behavioral definiens more precisely. For example, it might be thought that Laches could respond to Socrates' counterexample with a two-part definition of courage (one for hoplites, the other for cavalymen), young Charmides could put forth a multi-part definition of temperance (distinguishing those activities in which quietness or quickness is more temperate), and old Cephalus could amend his definition with an exception clause covering cases of madness, much like those found in legal contracts or insurance policies. But this is not how any of these exchanges in fact proceeds. Evidently, the conclusion drawn by Socrates and his interlocutor in each case is instead that no amount of such "tweaking" will save this sort of definition because there is something *fundamentally* wrong with all attempts to define the virtues in behavioral terms. As it happens, the discussions of behavioral definitions don't continue long enough to reveal the precise source of Socrates' dissatisfaction with them. Nonetheless, it is possible to speculate that it rests on the quite reasonable idea that even with the hypothetical amendments suggested above, an ingenious critic could simply construct new counterexamples to the more complex behavioral definitions. These new cases will then necessitate further disjunctions or exception-clauses, after which yet other counterexamples could be posed, and so forth, setting up an alternating dialectical pattern that could seemingly continue indefinitely.

It might appear that the advantage should go to the defense in this sort of situation, on the grounds that it gets the last move in any iteration of the repetitive



pattern. Clearly, Socrates would not agree. For if we extrapolate to the logical extreme of this hypothetical exchange, we will find the defense giving a long and complicated definition with a large number of disjunctive clauses, each specifying in *exhaustive* detail (1) a specific set of factual circumstances, and (2) the sort of behavior which would exemplify the defined virtue in those circumstances. Such a definition, the defense will maintain, might be unwieldy, but at least it would be *true*. Indeed, it might even be suggested that possession of such a finely textured behavioral definition would qualify as the sort of “precise” knowledge Socrates requires at *Euthyphro* 4e. Socrates’ stance is that whether it is true or not, it is not even a *definition*. In two prominent passages (*Euthyphro* 6c–d and *Meno* 72b–d) he argues forcefully against the acceptability of “enumerative” definitions on the ground that they fail to provide any illumination concerning the nature of the single entity *common* to all of the enumerated particular cases (*Meno* 74a) which *makes* them instances of the virtue in question (*Euthyphro* 6e). Thus, in the final analysis Socrates’ objection to behavioral definitions is itself disjunctive: each is either vulnerable to direct refutation by counterexample, or (being “enumerative”) is not a genuine definition at all.

In the end we seem to be left with a counterfactual conclusion with regard to the feasibility of using behavioral definitions to diagnose ethical behavior. *If* this sort of definition were appropriate, and if Socrates were somehow taught a correct definition of this sort, then it seems he *would* be able to use it to distinguish effectively between virtuous and non-virtuous actions. But of course, this conditional result is completely useless to Socrates’ program of practical ethics because he is opposed in principle to the first clause of the antecedent. Let us then consider whether the diagnostic function fares any better on Socrates’ own positive view about how the virtues should be defined.

### 1.3 Psychological Definitions and Ethical Diagnosis

I expect that the preceding sentence will provoke surprise and skepticism among some readers, who will protest that there *are* no such positive “Socratic” views about the nature of the virtues to be taken into account. After all, it will be objected, throughout his investigation of the virtues Socrates regularly and steadfastly denies that he himself has any knowledge, or even any positive beliefs, about the nature of the virtues, and it is never his own views, but always those of his interlocutors, that are on the table for examination and possible refutation. Moreover, it will be added, every one of the dialogues in which definitions of the virtues are sought ends quite conspicuously in *aporia* (a state of puzzlement), with Socrates and his interlocutors agreeing that the sought-after definition has eluded them, and worse, that they don’t know any more about the virtue in question now than they did at the outset.<sup>10</sup>

I don’t deny that this is an accurate representation of the “official” position presented in these dialogues. However, like many other recent writers,<sup>11</sup> I also believe that it would deny justice to Plato as a dramatist and a philosophical pedagogue to



settle for this superficial interpretation. For while it may be true that the views subjected to elenctic examination during these investigations are always officially those of Socrates' interlocutors, the progress of his interrogations nonetheless reveals significant information about his own positive views. In particular, careful attention to the sorts of questions Socrates poses, the sorts of objections he raises, and occasionally even positive suggestions he makes, indicates that he (or his author) is always steering—some might even say bullying—the interlocutors (or the reader) in a constant and definite direction with regard to the metaphysical question of what *sort* of thing the virtues are. I believe the ultimate source of Socrates' dissatisfaction with behavioral definitions is his recognition that even though patterns of overt behavior are publicly observable, and therefore well suited to diagnostic purposes, they have no rigid connections with the possession of virtues. The very same behavior pattern that in some circumstances manifests virtue may in other situations be non-virtuous, or for that matter vicious. In my view, Socrates' reaction to this observation is to reject behavioral definitions across the board, and to advocate in their place a completely different sort of definition that identifies each virtue with some distinctive state of an agent's soul. In the *Laches*, for example, after his earlier, behavioral definition as "staying one's post" is dispatched, Laches next proposes that courage is "a sort of endurance in the soul" (192b–c). Similarly, in the *Charmides*, at 162d–164d, Socrates uses very much the same sort of reasoning found at *Laches* 191a–b to convince his interlocutor (Critias) that the behavioral definitions he had lately been advocating in fact disguised his actual view, that temperance is self-knowledge (164d).<sup>12</sup> The clear advantage of such psychological definitions is an enhanced resistance to counterexamples, since the same state of one's soul can remain unchanged while producing a wide range of outward behavior in different circumstances.

Moreover, although the evidence for this is less direct, I believe that Socrates also advances a quite definite *intellectualist* view about what sort of psychological conditions comprise virtue, namely "knowledge of good and evil." Three different types of textual evidence support this attribution. To begin with, it is corroborated by the presence of two "intellectualist" corollaries (the unity-of-virtue thesis, and the denial of *akrasia*) in the non-definitional contexts of the *Protagoras*, where Socrates does seem to be defending a positive philosophical position. Second, the intellectualist conception of virtue is suggested by the manner in which Socrates refutes the various behavioral definitions discussed above. In each case, when he points to circumstances where a specified sort of behavior doesn't align well with the virtue in question, he may be taken also to be insinuating that someone who really had the virtue would know in what circumstances it *would* be appropriate to exhibit that or any other form of behavior. It is not difficult to see how "knowledge of goods and evils"—construed as the correct apprehension of all the positive and negative values attendant upon the outcomes of alternative courses of actions—would facilitate such optimal design of virtuous behavior.

The third and final piece of evidence that Socrates endorses this intellectualist conception of the virtues is provided by the order in which various types of definitions are taken up in these dialogues. We have already observed that in the

*Charmides* Socrates convinces Critias that what he really believes is that temperance is self-knowledge. But then, later on, at 174b–c, he rebukes Critias yet again in much the same manner by saying he now realizes this latter position *itself* masked the position Critias ultimately holds, namely that temperance is nothing but “knowledge of good and evil.” In a similar vein, the final definition of courage offered in the *Laches* (by Nicias at 194e–195a) is “knowledge of what is to be dared and what is to be dreaded,” which Socrates then goes on to argue is tantamount to “knowledge of all good and evil.”<sup>13</sup>

This intellectualist interpretation of Socrates’ positive position in the early dialogues is hardly novel. It has been endorsed in print so frequently over the past few decades that it may now fairly be regarded as a piece of orthodoxy.<sup>14</sup> Yet it has generally escaped notice that if Socrates does indeed believe that the virtues should be defined as inner psychological states, this seriously undermines the diagnostic function for the definitions of the virtues featured at *Euthyphro* 6e.

Let us consider first what becomes of the sufficiency thesis, which was identified above the most literal and natural interpretation of the passage. For the sake of argument, let us now suppose (1) that piety is correctly defined as knowledge of (certain) goods and evils,<sup>15</sup> (2) that Euthyphro knows this, and (3) that he has somehow been able to “teach” it to Socrates in whatever manner is required by *Euthyphro* 6e.<sup>16</sup> The important thing to notice at this point is that his possession of this definition would not enable Socrates to fulfill his stated intention at 6e, to diagnose Euthyphro’s prosecution or any other action simply by applying it directly to factual descriptions of these acts to determine whether they are pious or not. This, of course, is because the primary sort of definitions that Socrates favors are not virtuous *actions* but certain distinctive psychological states (more specifically, epistemic states) of virtuous persons. But unlike behavioral patterns, psychological states are not publicly observable. Thus, the very insight that fueled Socrates’ rejection of behavioral definitions in the first place, namely that the very same virtue (construed as a kind of knowledge) can give rise to a great variety of different sorts of behavior, also stands in the way of any straightforward application of Socrates’ preferred definitions to particular actions. This, then, is the very crux of the difficulty we are considering. Had it been possible to define the virtues in terms of overt behavior, the straightforward type of ethical diagnosis described at *Euthyphro* 6e could have gone through without much difficulty. It would simply have been a matter of determining whether the factual description of a given action matched the behavioral patterns specified in the definition. But as we saw above, Socrates is opposed in principle to that type of definition. On the other hand, on Socrates’ own intellectualist conception of the virtues, since (1) psychological states are not themselves publicly observable, and (2) there is no rigid link between the possession of a virtue and any particular form of behavior, it is hard to see how possession of a psychological definition he advocates could make effective diagnosis of ethical behavior possible. But this undercuts the reason Socrates offers at *Euthyphro* 6e for wanting definitions of the virtues in the first place. Now one could certainly imagine some more theoretically inclined philosopher taking consolation, and perhaps even vindication, from

the thought that even if these new definitions cannot perform this practical diagnostic function, they are nonetheless valuable insofar as they reveal the very essence of virtue. But that would be wholly out of character for our “single-minded moralist.”

But there is another possibility to be considered. What I just argued, strictly speaking, is that if there is any hope of using Socrates’ preferred definitions to diagnose virtuous actions, it will have to be accomplished in some more roundabout manner, and not in the straightforward way suggested by the language of *Euthyphro* 6e. Perhaps the diagnosis could be accomplished in two steps, by first determining that the agent of the act in question has the appropriate sort of “knowledge of good and evil,” and then determining that the act is one that flows from that knowledge in the appropriate fashion. Although the second step of this procedure is not unproblematic, the first step presents the major sticking point. The issue now is how a prospective ethical diagnostician could possibly certify that a given agent did possess the requisite knowledge of good and evil, and more generally how anyone can determine that another person possesses any sort of knowledge whatsoever.

As it happens, this more general epistemological issue arises in a passage near the end of the *Charmides* (170b–172a) where Socrates is evaluating the proposed definition of temperance (*sôphrosynê*) as “self-knowledge.” Rather than construing this expression in the most natural manner as “knowing oneself,” Socrates interprets it somewhat perversely as equivalent to “knowledge of knowledge (and ignorance),” and then proceeds to consider whether it is humanly possible to achieve this sort of meta-knowledge. In the course of this, he introduces a distinction between two different ways in which it might be thought possible for someone to assess the epistemic states of others. On one of these alternatives, the person making the assessment knows only *that* some claimant or other does or does not actually possess some knowledge, but does not possess that first-order knowledge herself, and on the other, she knows *what* is claimed (truly or falsely) to be known by someone else, and presumably can use this as a benchmark to determine whether a claim to possess that first-order knowledge is genuine or spurious. The question before us now is whether either of the conceptions of “knowledge of knowledge” surveyed in the *Charmides* can be of use to Socrates in using his intellectualist definition of the virtues in the two-step form of ethical diagnosis under consideration.

It is quite possible that the presence of the first conception of meta-knowledge in the *Charmides* amounts to a veiled reference to Socrates’ own method, since he is represented throughout the early dialogues both disclaiming ethical knowledge himself, and at the same time deploying his elenctic skills to undercut claims to ethical knowledge made by his various interlocutors. If this is the case, it should be kept in mind that this critical method is much better suited to demonstrating, through repeated refutations, that a given interlocutor *lacks* knowledge than it is to establishing that someone really *possesses* it. There is certainly no record of cases of elenctic “certification” anywhere in the early dialogues,<sup>17</sup> and it very difficult to imagine one of Socrates’ elenctic encounters producing such a positive result. Consequently, it is highly doubtful that this sort of “knowledge of knowledge” could enable Socrates to distinguish effectively between virtuous and non-virtuous action as promised at *Euthyphro* 6e.

Let us now consider whether the more robust conception of “knowledge of knowledge” described in the *Charmides* can play a fruitful role in the use of psychological definitions to perform ethical diagnoses. Generally speaking, it is one thing to know what a given property amounts to, and quite another to possess that very property. What is more, this distinction holds even in cases where the property in question is the exemplification of some cognitive state. I might, for example, know that to be a genius is to have an IQ score of at least 175 without having achieved that score myself, and conversely, there are many people with scores that high who remain entirely oblivious to the concept of genius. As this pertains to our present concerns, it means that (1) knowing that a virtue is a certain type of knowledge, and (2) possessing that very type of knowledge, are independent of one another. Still, we can imagine a prospective ethical diagnostician who has been taught the Socratic definition of, say, piety, and who *also* just happens to satisfy that very definition, and so possesses the requisite sort of knowledge of good and evil. The suggestion we are considering, then, is that such a person might somehow be able to use the *content* of this knowledge as a “model” in determining whether some agent also possesses that same knowledge, and ultimately whether the actions of that agent are pious or not. Two separate textual considerations weigh heavily against this suggestion. In the first place, it should be simply recalled that in the *Euthyphro* passage it is not some hypothetical diagnostician but Socrates himself who proposes to make use of the definition he demands from his interlocutor, and any suggestion that *he* stands ready to conduct a “privileged” sort of ethical diagnosis available only to *virtuous* diagnosticians clashes with his repeated denials throughout the early dialogues that he himself possesses wisdom, or by implication, any of the virtues.<sup>18</sup>

Of course, negative universals are notoriously difficult to prove, and my principal contention here is not that there would be no possible way for Socrates to use one of his own preferred intellectualist definitions to diagnose virtuous and vicious actions, as he claims to want to do at *Euthyphro* 6e. My point is rather that the sort of indirect diagnosis just described requires such substantial amount of crucial information about the epistemic condition of the agent that it would be highly misleading for Socrates to cite the sufficiency thesis as his controlling reason for wanting Euthyphro to teach him the nature of piety. In other words, if he means to say that if Euthyphro teaches him the definition *and* he also just happens to satisfy that definition, then he will be able to diagnose ethical action, it would be grossly misleading of him to declare that the possessing the definition by itself will suffice for this purpose. Yet as we saw above, that is exactly what Socrates appears to be saying on the most natural interpretation of *Euthyphro* 6e.

Furthermore, it is not just the sufficiency thesis that becomes suspect on the Socratic conception of the virtues. If, as was argued above, Socrates is indeed quietly advocating intellectualist definitions of the virtues, this also provides good reason to wonder about his apparent subscription at *Euthyphro* 4e to what was referred to above as the *necessity thesis*, that it would be impossible to diagnose pious and impious actions *without* knowing the definition of piety. Making use of the distinction introduced above, let us imagine now that there is someone who *satisfies* this definition, but for some reason doesn't *know* the definition,<sup>19</sup> perhaps because this person

had never been exposed to Socrates, whom Aristotle describes as “the first to search for universal definitions [of the ethical virtues]” (*Metaphysics* M.4 1078b17–18).<sup>20</sup>

Let us also assume, as before, the truth of the “intellectualist” definition of piety as “knowledge of (certain) goods and evils.” By hypothesis, such a person would have the ability to discern the goods and evils inherent in the outcomes of *every* action (whether their own or those of someone else), and if this sufficed to enable him to choose the correct action for himself on every occasion of choice, it is difficult to see why it would not also enable him to determine whether any past or future action of *anyone* is virtuous or not. Granted, never having been concerned with the investigation of the virtues as a *theoretical* pursuit, the person we are imagining could not be said to possess the *concept* of virtue (or any particular virtue). Indeed, she might not even be familiar with the *name* of the virtue, and consequently might not satisfy the technical requirement of *Euthyphro* 6e of being able to *say* which actions are pious and which are not. But for a “single-minded moralist” like Socrates, that would seem to be no reason to deny that person the ability to perform the sort of ethical diagnosis described in the passage.<sup>21</sup>

## 1.4 Mitigation

Let us now consider the range of open interpretive possibilities on the supposition that the sufficiency and necessity theses are both subverted by Socrates’ own intellectualist conception of the virtues. If, to begin with, we consider the matter from *within* Plato’s dramatic framework and focus just on his protagonist, one possibility is that Socrates is aware that problems arise with the diagnostic function on his own conception of virtue, but insists upon it anyway at *Euthyphro* 6e merely as a dialectical ploy to lure Euthyphro into the elenchus. Alternatively, we might speculate that Socrates himself just doesn’t realize that his original rationale for requesting these definitions will go by the boards if his own views about the nature of the virtues are adopted.

But of course, if *we* can contemplate these possibilities, so too might have Plato, and what he might have contemplated he might have created. Thus, we now seem to face four equally unappealing interpretational options. If, on one hand, we suppose that Plato maintains an authorial distance from his protagonist, we might conclude that he intentionally portrays Socrates as deceptive or confused in one of the ways just described. On the other hand, if we suppose that the figure of Socrates is simply a faithful mouthpiece for the views of his author, we are left to infer either that Plato for some reason intends to conceal from his readers the problems uncovered above, or that there is incoherence and confusion in Plato’s own thinking about the program of defining the virtues.<sup>22</sup>

I do not believe it is possible in the end to avoid a difficult choice among these possibilities, but I do think that something can be done at least to soften the edge of the difficulty I have been expounding. For even though I’ve argued that possession of Socrates’ preferred intellectualist definitions of the virtues would be neither

sufficient nor necessary for the successful diagnosis of virtuous action, this does not lead automatically to what would be a much more damaging conclusion, that the possession of such definitions could have no practical ethical use *whatsoever*, which would place Socrates' pursuit of these definitions entirely at odds with the pronounced practical bent of the earlier dialogues. In fact, quite to the contrary, I believe that a reasonable case can be made that even a "single-minded moralist" like Socrates would have considerable use for such definitions.

In the first place, it is not even clear yet that Socrates' intended diagnostic function for definitions is a lost cause. Even though it was argued that the "epistemic" conception of the virtues undermines both the sufficiency and necessity theses, it remains to be considered whether Socrates could fall back to the weaker *utility* thesis by holding that possessing a definition could contribute in some way or other to the ethical diagnosis of action, even if it was neither necessary nor sufficient for that end.<sup>23</sup> In fact, it is possible to see how this could be so by considering the following hypothetical case in which Socrates himself might employ his unmatched prowess in the elenchus to diagnose the very action that initially gave rise to these issues in the *Euthyphro*. Let us then imagine once more that Socrates was somehow<sup>24</sup> able to discover what we are supposing is the correct definition of piety, "knowledge of (certain) goods and evils," and that he has set himself to diagnosing Euthyphro's dubious prosecution of his own father.

Notice that even prior to commencing the diagnosis, Socrates knows, on the basis of the definition, that *if* the prosecution were a pious act, it would follow that Euthyphro does in fact possess the specified sort of knowledge, and also that his action results from his possession of this knowledge. As it happens, however, Socrates has at his disposal a superbly designed instrument for testing the truth of the first of these conditions. I remarked earlier that it is quite doubtful that the Socratic elenctic method could ever serve to "certify" that someone really does possess knowledge, but is extremely well suited to establishing that someone, such as Euthyphro, who claims to have knowledge, in fact *lacks* it. And indeed, Socrates' repeated elenctic refutations throughout the entire remainder of the dialogue demonstrate beyond any doubt that Euthyphro does not have the "precise knowledge of things divine" which he claims for himself at 4e. But since Euthyphro's decision to prosecute his father could not have been made on the basis of knowledge that he does not possess, Socrates' elenctic treatment of Euthyphro conclusively establishes, on the "intellectualist" conception of piety, that the prosecution could not possibly be a pious act.<sup>25</sup>

Finally, from a more general perspective, there is another quite different respect in which Socrates' tendency to conceive of the virtues as intellectual achievements fits well with his single-minded moralism. For it would be a mistake to suppose that his practical reasons for investigating the virtues stem exclusively from a hope or belief that the results of these inquiries will be helpful in *recognizing* virtuous and vicious action. To the contrary, a number of strategically placed passages in the early dialogues indicate quite clearly that as important as he thinks ethical diagnosis is, he is at least interested, and perhaps even more interested, in the question of how virtue might be *imparted*, especially to the young. This in fact is the very first



substantive topic taken up in both the *Laches* and the *Protagoras*, and in both these cases Socrates explicitly justifies raising subsequent questions about the *nature* of virtues on the grounds that they are necessary preliminaries to figuring out how they might be instilled.<sup>26</sup>

Now, it might seem that Socrates' intellectualist conception of virtue would lead him directly to the view that imparting virtue will simply be a matter of *teaching* this sort of knowledge (as opposed, say, to some program of behavior modification<sup>27</sup>). In fact, however, some quick biographical research into the great figures of Athenian history (who were reputedly virtuous, but conspicuously failed to raise virtuous sons) leads Socrates to conclude that this gives rise to a well-known Socratic paradox, namely that virtue appears to be a very peculiar sort of knowledge that *cannot* be taught.<sup>28</sup> It is well known that the force of this paradox ultimately causes Plato to initiate new inquiries into the *non-ethical* concepts of knowledge and learning, and that his pursuit of these and related topics lead him through the epistemological thickets of the last section of the *Meno*, and eventually onto the lofty metaphysical terrain of *Republic* V-VII. But this, of course, is a trail that a "single-minded moralist" could not follow.

## Notes

1. I will be working here within a moderate, "coarse-grained" developmental interpretation of Plato's works according to which there is significant doctrinal difference (and even incompatibility) between one group of early, or "Socratic," dialogues on one hand, and another "middle period" group (which includes the *Republic*) on the other. In particular, the dialogues are assumed here to fall within the "Socratic" group are *Apology*, *Crito*, *Euthyphro*, *Laches*, *Charmides*, *Protagoras*, *Gorgias*, and *Republic* I (on which see n. 12 below).

I will, however, systematically avoid the issue of whether the doctrines of the "Socratic" dialogues can or should be attributed to the historical Socrates (as opposed to being views held by the young Plato, or for that matter, simply hypothetical positions never held by Plato or Socrates). Accordingly, unless otherwise indicated the intended referent of "Socrates" below will always be Plato's protagonist rather than his mentor.

2. Vlastos (1991, p. 62 n. 68).
3. It is plausible to see this report as elaborating upon *Apology* 26d–e, where Socrates emphatically disassociates himself from the concerns of the natural philosophers (*hoi physikoi*), and singles out Anaxagoras for special mention. For even though the *Phaedo* is generally regarded as a "middle period" work, I believe it is nonetheless reasonable to assume that Plato would naturally have intended the "autobiographical" details related there to cohere with (and fill in) those given in the earlier work. Having a single character express different (and even incompatible) philosophical views at different times is perhaps understandable if Plato wants to retain his protagonist as his own philosophical positions shift; giving him alternative life-histories at different times would just be perverse. In accordance with the methodological remarks in note 1, I will not be at all concerned here with the issue of whether, or to what extent, these "autobiographical" passages describe Socrates' actual intellectual development, as opposed to presenting a fabricated life-history of a fictional character simply for purposes of dramatic vividness.
4. Essentially the same point is made by Brickhouse and Smith (1994, pp. 62–3) and H. Benson (2000, p. 145).
5. Plainly, such factual descriptions cannot be limited to purely physical descriptions of the acts in question, but must also include at least some information about S's deliberative structure,

since it is arguable that one can't really know what a given action is without taking into account such things as the agent's own view about what she is doing, and her reasons for doing it. Indeed, this thought is reinforced by the passage in the *Phaedo* where Socrates describes the moment of his disillusionment with Anaxagoras:

It was a wonderful hope, my friend, but it was quickly dashed. As I continued to read I discovered that [Anaxagoras] made no use of mind, and assigned to it no causality for the order of the world, but instead adduced causes like air and ether and water and many other bizarre things. It seemed to me that he was just about as illogical as if someone were to try to account in this manner for my conversing with you, by adducing causes such as sound and air and hearing and a thousand others, and never bothered to mention the real reasons, which are that Athens has thought it better to condemn me, and that I, for my part, have therefore thought it better to sit here, and more right to stay and submit to whatever penalty she orders. (*Phaedo* 98b–e)

(Interestingly, this passage may be an ancient pre-adumbration of the distinction between the “physical stance” and the “intentional stance” as alternative manners of describing human behavior in Dennett (1978).)

At the same time, for reasons that will emerge below, it would be a mistake in the opposite direction to build too much into these factual descriptions by requiring exhaustive information about the agent's deliberative structure, in particular information pertaining to the settled dispositions that make up one's character. Such a liberal approach to describing actions ultimately leads into the Aristotelian doctrine that virtuous actions must be defined in terms of virtuous character (e.g., as acts that a virtuous person would do while acting in character). Since I believe this is contrary in spirit to Socrates' fundamental idea that the definition of piety will enable or help him to diagnose actions *directly* (without having to first assess the character of the agent), I will adopt a comparatively “thin,” conservative understanding of what Socrates would allow in the factual descriptions of actions needed for his diagnostic purposes.

6. See, e.g., Geach (1966), Santas (1972), Burnyeat (1977), Irwin (1979), Nehamas (1986), Berverisluis (1987), Vlastos (1990), Brickhouse and Smith (1994). The alleged fallacy is a circularity that is supposed to come out of (1) Socrates' commitment to this epistemological principle, (Nec) If S lacks a definition of F, S cannot pick out examples of F, and (2) his simultaneous use of examples to *test* proposed definitions. Socrates' defenders on this issue typically proceed by advancing weaker interpretations of (Nec) which are obtained by understanding what is meant by the “picking out” of examples in the consequent quite strongly with respect to *scope*, the *level of justification* involved, or both. The range of possible interpretations of the principle can accordingly be seen as delimited by the following two extremes.

Strongest (Nec) If S lacks a definition of F, S cannot form a reasonable true belief that *any* case is an example of F.

Weakest (Nec) If S lacks a definition of F, S cannot *know* of *every* case whether it is an example of F.

This interpretational issue will have no material bearing here. In Section 1.3 I shall identify a difficulty with the necessity thesis that arises for *any* of these versions, down to and including Weakest (Nec).

7. Cf. Benson (2000, p. 157).
8. Other scholars have also perceived this apparent difficulty with the sufficiency thesis, and have argued in different ways that it is merely apparent. Kraut (1984), for example, argues that for Socrates knowing “what a virtue is” entails nothing less than “having an entire ethical theory.” Against this, Brickhouse and Smith (1994) correctly argue that the texts simply give no indication that in asking “what the virtues are,” Socrates is looking for anything more than correct *definitions* of the virtues (pp. 61–4). Brickhouse and Smith themselves argue that the obvious implausibility of the sufficiency thesis provides good reason *not* to attribute it to Socrates (pp. 62–3). I am on the whole sympathetic with their attempt, since I believe that



the sufficiency thesis is ultimately incompatible with aspects of Socratic intellectualism, but I just don't see how the relatively plain language of *Euthyphro* 6e will permit withholding this attribution, and I am not persuaded by their attempts to show otherwise. In particular, the case they make (pp. 62–3) against literal sufficiency and in favor of what I am calling “sufficiency in the circumstances,” while correct, is also unhelpful, since the difficulty arises even with the qualification.

9. On the conjecture that *Republic* I was originally a freestanding “Socratic” definitional dialogue to which the remainder of the *Republic* was appended during the “middle period,” see Irwin (1995, p. 376 n. 1).
10. *Euthyphro* 15c–16a, *Laches* 159e, *Charmides* 175a–176e, *Republic* I 354b–c.
11. See especially Burnyeat (1971) and Santas (1971), but also Ferejohn (1982), and Brickhouse and Smith (2000).
12. The situation in *Republic* I is somewhat more complicated, but is it at least possible to see the same sort of movement away from “behavioral” definitions of justice given by Cephalus (“not cheating and paying one’s debts”; 331b) and Polemarchus (“helping one’s friends and harming one’s enemies”; 332d). Whether a movement towards psychological definitions occurs in Book I or is delayed until Book IV is less clear, but it may be implicit in Thrasymachus’ definition of justice as “what is in the interest of the strong ruler” (339a). If this is taken to mean what is *really* in the ruler’s interest, it is arguable that behaving justly would indeed require knowledge of what’s good and evil (for the ruler).
13. Like other writers who favor this sort of interpretation, I take the fact that none of these later definitions escapes refutation to be indicative of Plato’s dramatic and pedagogical objectives, and not to be decisive as to whether the definitions are ultimately regarded as unacceptable by either the character or the author. In the case of the *Laches*, even though the conclusion that courage, on the proposed definition, would be identical to the whole of virtue might reasonably be expected to induce *aporia* in Socrates’ interlocutors, it should not necessarily be taken as a decisive *reductio*. In fact, the evidence of the *Protagoras* indicates that this “aporetic” result, the unity-of-virtue thesis, is in fact Socrates’ own considered view. In addition to this, it is significant that Nicias apparently gives credit to Socrates himself as the source of his proposed definition (*Laches* 194c–e), an attribution which is corroborated at *Protagoras* 359c–e, where Socrates seems to be using essentially the same definition as a dialectical premise.
14. See note 11 above.
15. But which goods and evils in particular? Following much the same lines as our earlier discussion of *Republic* I in note 12, one might plausibly take Euthyphro’s final definition in the dialogue, “service to the gods,” as importing the intellectualist conception of virtue on the grounds that such “service,” if it is to be anything more than acceding to the gods’ *commands*, will require prior of knowledge of what is *in fact* in their interest, i.e., the goods and evils that will come to them as the result of various possible human ministrations.
16. Presumably, this will involve more than Euthyphro just *telling* Socrates the definition and Socrates believing him. Otherwise Socrates would be open to the charge of disingenuousness, since, according to the arguments just given, Socrates would be asking for something he already possesses. I suspect, therefore, that in this and similar contexts Socrates will not regard a definition as properly “taught” until it has been proposed and adequately *defended* at some length in the elenchus. This naturally raises the more general interpretive question of why Plato, through Socrates’ regular professions of ignorance, structures the dialectics of these investigations in such a way that if the correct (Socratic) definition of a virtue *were* ever proposed explicitly, it would have to be introduced by one of the interlocutors, with Socrates himself taking the role of elenctic examiner. One plausible answer is that Plato recognizes that the degree of epistemic warrant earned by sustained elenctic survival depends significantly on the quality and severity of the elenctic challenge, and holds up Socrates as the most adept practitioner of the elenchus.
17. The only possible, though highly unlikely, exception is a passage in the *Gorgias* where Socrates himself withstands a brief and feeble attempt by Polus to refute him elenctically before reassuming his usual role as questioner (465e–466d).

18. See Benson (2000, pp. 238–43). The generalization follows immediately if it is assumed that Socrates consistently endorses the unity-of-virtue thesis (see *Protagoras* 329a–333e). Moreover, there is independent evidence that he at least believes he lacks temperance in a comic passage at *Charmides* 155c–e where he confesses his inability to resist taking a peek inside his young interlocutor’s cloak.

In the face of such evidence, a defender of the view that Socrates is nonetheless virtuous would have to maintain that his numerous professions to the contrary are either deceptive or mistaken. I shall not pause over the first option, since I believe that ascriptions of irony or insincerity could in principle be used to rescue *any* interpretation from textual refutation, and should therefore be avoided except as a last resort. The second option, that Socrates (or anyone else) could be “unwittingly virtuous” is more interesting, and will be taken up below.

19. Here again we encounter the case of the “unwittingly virtuous” person introduced in the preceding note. This possibility in effect challenges the thesis, which is hardly self-evident in any case, that anyone who possessed a virtue would necessarily be aware of its presence within, and would as a consequence be sufficiently acquainted with its nature to be able to provide a definition for it. As it happens, Socrates himself might be taken to endorse some such “transparency” thesis at *Charmides* 158e–159a where, after brushing aside Charmides’ modest denial that he possesses temperance as a predictable manifestation of that very virtue, he then forces the boy into proposing a definition of the virtue by means of the following reasoning:

If temperance abides in you, you must have an opinion about her. She must give some intimation of her nature and qualities, which should enable you to form some notion of her.

However, while this passage could arguably be taken to express the implausibly strong transparency thesis formulated above, according to which any one who possesses virtue must be aware of it and be able to say what it is *immediately*, it might also be interpreted more conservatively as making the weaker point that such a person, because of the virtue’s presence, should *eventually* be able to recognize and define it, although the process towards this end might be long and difficult, and might perhaps even require some “maietic” assistance (cf. *Theaetetus* 150b–151d, 157c–d).

20. Since the attribution is based upon Aristotelian history of philosophy, the intended reference of “Socrates” in this passage is apparently the historical figure and not the quasi-fictional protagonist of the Platonic dialogues.
21. Notice also that the ability to say *which* virtue a given action exemplifies loses importance in light of Socrates’ commitment to the unity-of-virtue thesis.
22. For the sake of fairness and completeness, there is a fifth interpretational possibility that should be mentioned. Someone predisposed to resist the entire intellectualist line of interpretation advanced here would no doubt view my arguments here as a *reductio* of the view that Socrates advocates an “epistemic” conception of the virtues throughout the definitional dialogues. A thorough evaluation of this alternative would take us too far afield here, but it will ultimately turn on the issue of whether its proponents are able to make a persuasive case that the evidence adduced to document the presence of the “epistemic” conception of the virtues in the definitional dialogues can be interpreted otherwise.
23. As mentioned above in note 8, Brickhouse and Smith (1994) evidently believe that *Euthyphro* 6e commits Socrates to nothing stronger than the utility thesis (pp. 62–3). By contrast, my position is that the most natural interpretation of the passage commits him to the sufficiency thesis, but that it would be reasonable for him to mitigate the difficulties discussed here by falling back to the utility thesis in order to retain the diagnostic function for his “epistemic” definitions of the virtues.
24. Of course, given what transpires in the course of the dialogue, the means of that acquisition plainly could not possibly be Euthyphro’s tutelage.

25. This would not, of course, be sufficient to establish that it was an *impious* act, since conclusive positive diagnoses of impiety also would presumably require the “precise” knowledge of piety that Euthyphro demonstrably lacks and Socrates himself never claims to possess.
- It might perhaps still be open to Euthyphro to argue that his act, even though it has been shown not to be pious, might nevertheless be “pious-like” if it mimicked an act that a pious agent *would* perform in similar circumstances. But since the dialogue establishes beyond question that Euthyphro doesn’t have *any* defensible definition of piety, much less what we are supposing is the correct “intellectualist” one, it follows that in addition to being unable to choose pious *actions*, he also lacks any ability to diagnose pious *persons*, which means that he would also be incapable of selecting the right person to mimic. Hence, even if an action of his happened to emulate that of a genuinely pious agent, it would have to do so by dumb luck, a possibility that may be safely ignored.
26. Cf. also the very first lines of the *Meno* (70a), which pose the question of whether virtue can be taught. Even though this dialogue is generally thought to be a later composition written when Plato was in the midst of a critical reconsideration of the doctrines of the early dialogues, it also appears that the first third of the work (through 79e) appears to be a deliberate recreation by Plato of a typical Socratic elenctic inquiry, and may therefore be plausibly taken to represent the philosophical program of the earlier dialogues.
27. Some such program may be intimated in Aristotle’s *Nicomachean Ethics* II 2–3.
28. *Protagoras* 319e–320c (and cf. *Meno* 93a–96e).

## Bibliography

- Benson, H. 2000. *Socratic Wisdom*. Oxford: Oxford University Press.
- Berversluis, J. 1987. “Does Socrates Commit the Socratic Fallacy?” *American Philosophical Quarterly* 24:211–23.
- Brickhouse, T., and N. Smith. 1994. *Plato’s Socrates*. Oxford: Oxford University Press.
- Brickhouse, T., and N. Smith. 2000. *The Philosophy of Socrates*. Nashville, TN: Westview.
- Burnyeat, M. 1971. “Virtues in Action.” In edited by G. Vlastos.
- Burnyeat, M. 1977. “Examples in Epistemology: Socrates, Theaetetus, and G.E. Moore.” *Philosophy* 52:381–98.
- Dennett, D. 1978. “Intentional Systems.” In edited by *Brainstorms*, D. Dennett. Denver, CO: Bradford.
- Devereux, D. 1977. “Courage and Wisdom in Plato’s *Laches*.” *Journal of the History of Philosophy* 15:129–41.
- Ferejohn, M. 1982. “The Unity of Virtue and the Objects of Socratic Inquiry.” *Journal of the History of Philosophy* 20:1–21.
- Geach, P. 1966. “Plato’s *Euthyphro*: Analysis and Commentary.” *Monist* 50:369–82.
- Irwin, T. 1979. *Plato: Gorgias*. Oxford: Oxford University Press.
- Irwin, T. 1995. *Plato’s Ethics*. Oxford: Oxford University Press.
- Kraut, R. 1984. *Socrates and the State*. Princeton: Princeton University Press.
- Nehamas, A. 1986. “Socratic Intellectualism.” *Proceedings of the Boston Area Colloquium in Ancient Philosophy* 2:275–316.
- Penner, T. 1973. “The Unity of Virtue.” *Philosophical Review* 82:35–68.
- Santas, G. 1971. “Socrates at Work on Virtue and Knowledge in Plato’s *Laches*.” In edited by G. Vlastos.
- Santas, G. 1972. “The Socratic Fallacy.” *Journal of the History of Philosophy* 10:127–41.
- Vlastos, G., ed. 1971. *The Philosophy of Socrates: Critical Essays*. New York: Anchor Books.
- Vlastos, G. 1990. “Is the Socratic Fallacy Socratic?” *Ancient Philosophy* 10:1–16.
- Vlastos, G. 1991. *Socrates: Ironist and Moral Philosopher*. Ithaca: Cornell University Press.

# Chapter 2

## Definition and Elenchus

Nicholas White

### 2.1 The Supposed Problem of the Elenchus

Early in the 1980s Gregory Vlastos called attention to what he took to be a problem afflicting the thinking of the character Socrates in Plato's early works.<sup>1</sup>

Since that thinking is commonly associated with the word "elenchus" (which scholars now treat, thanks to Latin, as an English word, since the translation "refutation" isn't quite true to the Greek "*elenchos*") the problem has come to be called "the problem of the elenchus"—though for reasons that will become clear I think that it isn't really a problem. Rather it was brought into existence by the particular way of thinking about socratic philosophical activity that Vlastos adopted.<sup>2</sup>

In this brief study I'll concentrate on the instances of elenchus that have to do directly with the scrutiny of definitions.<sup>3</sup> These form a well-defined class with their own systematic methodological features and consequences.

To be sure, by no means all elenchi are directed against definitions. Some are directed against theses that are similar to definitions in being identities, such as the famous thesis that virtue is knowledge (*Prt.* 361a–d), which are closely linked to attempts to define virtue. Other targets of elenchus are different, for example *non-identities*, such as the equally famous thesis that virtue *isn't* knowledge.

Although a more extended treatment of elenchus would have to embrace non-definitional targets of elenchus, nevertheless searches for definitions and refutations of candidates form such a salient and philosophically significant feature of the socratic landscape that the elenchi that are concerned with definitions require to be treated together in their own right.<sup>4</sup>

Taken as applied to definitions, the problem of the elenchus arises from the following fact.<sup>5</sup> Although at the end of a standard elenctic argument, Socrates claims that a candidate definition put forward by an interlocutor has been refuted, the most that can be claimed, according to Vlastos' view, is that the proposal is inconsistent with a number of other assumptions made by the interlocutor, which are used as

---

N. White (✉)  
University of Utah, Salt Lake City, UT, USA  
e-mail: nicholas.white@philosophy.utah.edu

premises in the supposed refutation. According to this view, the assumptions used against the proposed definition are just that, assumptions, with no more claim to credence than the proposal itself.

So even if it's shown that the assumptions and the proposal are mutually inconsistent, Vlastos contends, all that we're entitled to conclude is that *something* in this inconsistent set has to be rejected—but not necessarily that the proposed definition itself is that something.<sup>6</sup> The question then is, how can it ever be claimed that the proposed definition is refuted? If this demonstration of inconsistency were all there is to the elenchus, there simply couldn't *be* any refutations of proposed definitions in the socratic dialogues. But the dialogues themselves purport to present plenty of such refutations. That's the problem.

In order to respond to the problem, Vlastos urges that we must show why a socratic elenchus refutes the candidate definition rather than the judgments used as premises against them.<sup>7</sup> We thus have to show a kind of asymmetry between the candidate and the premises. In order to do this properly, Vlastos contends (and this is where his mistake begins), we have to show why Socrates takes himself generally to be entitled to take the assumptions, rather than the candidate, to be *true*. Otherwise, Vlastos thinks, we wouldn't have a refutation of the requisite sort. To make this showing of truth, Vlastos also thinks (1994, p. 19) we have to introduce an epistemological element. It's not enough just to *call* the premises true. They have to have some epistemological status that the candidate definition doesn't have, and that justifies that ascription of truth. For what's required is that at the end of the whole elenchus, something has to be “proved” (*apodedeiktai*) true or false.

## 2.2 Do the Premises Have To Be True?

Although the overall framework that Vlastos sets up for the treatment of this matter seems to me to be misconceived, some parts of it are clearly right. He's right that, given the inconsistency between candidate definitions and judgments used against them, there must be some account of why an elenchus refutes the former rather than the latter. He's also right to assert that in these works, Socrates is portrayed as attaching great value to truth (1994, pp. 5–7) and as himself striving to attain it. In addition, Vlastos is correct that Socrates isn't pictured as offering any general method governing his investigations.<sup>8</sup>

On the other hand Vlastos is mistaken in believing that truth must in the end be attained by showing that *the aforementioned premises* are true. He's also mistaken in taking the elenctic procedure to require the premises to have the epistemological status that he claims for it. In fact the elenchus, as it's conceived in these works, and for the purposes that it must serve when directed against a definition, doesn't require any such thing.

It may seem strange to have it said that a refutation can be successful even if it's not somehow settled that the premises used in it are true. The strangeness, however, is only apparent. Once one understands what the socratic definitional project is, and

what a definition is expected to do, and what it therefore takes to show a definition to be inadequate, the sense of strangeness vanishes.

I want to begin to show this by considering a proposal made by Hugh Benson (2000) in his own argument that Vlastos' problem doesn't in fact arise (see, e.g., p. 92). In response to Vlastos, Benson holds that socratic refutations of definitions can be successful. For Benson contends that all that's required for the success that Socrates is after is merely to show an inconsistency between the proposed definition and other judgments that the interlocutor in fact holds. No need, Benson says, to make an additional effort to show that those judgments are true, or even to show why an elenchus refutes the proposed definition rather than the judgments. This proposal sets the bar where an elenchus can clear it.

Benson's explanation of the success of elenchus is that Socrates' aim in it is simply to expose the ignorance of his interlocutors. Showing that an interlocutor holds inconsistent judgments is enough for that. If a set of your judgments is inconsistent, then at least one of them must be false. But if something that you accept is false, then you don't know it; and if it's a part of the conjunctive basis of your pretensions to knowledge, then your claim to knowledge is false too—without our having to show, after all, *which* of your judgments is false.

In an important way I think this view of Benson's is right, but not entirely. Socrates isn't out, I think, *merely* to show that his interlocutors are ignorant. In my opinion he's out to show that they're ignorant *in a particular way*: they're ignorant of how to *define* certain terms that they (and Socrates) wish to use in important judgments which have practical import. There are plenty of ways of trying to show that someone's ignorant. Some of these ways are used, in fact, in the dialogues that don't focus on defining a particular term (e.g., the *Gorgias* and the *Euthydemus*). We need an explanation of why the interlocutors in the definition-seeking dialogues are argued to be ignorant in this special definition-focused way.

The socratic dialogues concerned with definitions all *say*, at their conclusions, that proposed definitions are refuted. But not just that. In addition, the further actions of the participants and the course of the discussions show that those refutations have decisive consequences. Except when a refuted definition is reformulated so that a refutation no longer applies to it (e.g., *Rep.* 331d–e), the old version of the definition is no longer treated as a live option. It's dropped. Thenceforth a new definition is sought.

Furthermore, when all candidate definitions that present themselves are refuted, the investigations for whose sake a definition was sought are expressly treated as stymied. They have to be suspended until an adequate definition be found. Definitions aren't sought for their own sake, but for the sake of dealing with further questions.

These considerations seem to me to show that Socrates' goal here can't be simply to demonstrate his interlocutors' ignorance. It must have something more directly to do with the business of defining—what it requires, how it can fail, and what consequences its doing so has. Otherwise the refutation of definitions wouldn't play such a large role in these works.

Nevertheless Benson's suggestion does point to an important fact. It's that Socrates may be able to accomplish his elenctic purposes, and indeed in the end attain the goal of truth, without giving a reason for holding that the premises of the elenchus used against candidate definitions have some stronger epistemological status, or some stronger claim to truth, than the candidates themselves. That turns out to be the crucial thing to see.

### 2.3 An Elenchus is No Cartesian Meditation

I want now to make a detour, to discuss a factor that makes contemporary readers prone to misjudging the structure of the elenctic project. That factor is a tendency to assimilate socratic concerns to those of contemporary epistemology and to some Cartesian or quasi-Cartesian project that generated them. I don't in the least deny that Socrates is portrayed as devoted to discovering truth. The issue, however, concerns the route by which he hopes to reach it. Nor am I one of those who deny that the socratic dialogues are concerned with epistemology, though there are many open questions as to just what form it takes in them.<sup>9</sup> I do think, however, that something essential is lost when one doesn't interpret these epistemological concerns within the right philosophical context. The philosophical context that's relevant here is created by the role that Socrates is made to think of definitions as playing.

Many recent interpreters, including Vlastos, have been wrongly influenced by, among other things, various socratic professions of ignorance, and also by what the interpreters take to be evidence that according to these dialogues, a person can't rightly accept any statement involving a term until a definition of the term has been established.<sup>10</sup> Such interpreters think that throughout the socratic conversations, *any* judgment is under suspicion prior to the defining of its central term (however that might be picked out), and that the judgments need to be justified by such a definition before it can be asserted. But since the dialogues begin with no established definitions, that means that all such judgments in them start out under a cloud.

Accordingly these interpreters come gradually to think of this socratic project as somewhat akin to a more or less cartesian one, with all or nearly all judgments subject to doubt at the start, until some of them are cleared of it by some method that the project devises and applies. Thus the judgments used as premises in refutations of proposed definitions seem to be just as much in need of scrutiny as the proposed definitions themselves.

In this way the idea that definitions ever really can be refuted collapses—for want of any firm premises to use in the refutation.<sup>11</sup> On these terms these dialogues can't contain any refutations. Something then has to be done, as Vlastos saw, to rescue the textually unambiguous fact that in those dialogues many definitions are indeed said to be refuted and are disposed of.

Indeed, something has to be done to rescue not merely the fact that the dialogues say that they refute definitions, but the fact that they *even undertake to scrutinize them at all!* This thought was crystallized in an article by Peter Geach in the mid-1960s, under the label "the socratic fallacy." This supposed fallacy would arise



from thinking that no judgment involving a term can be accepted until a definition of the term is established, and that a definition can be established only by testing it against independently accepted judgments involving the term.

The former of these two conjuncts would rule out, right from the start, the testing enjoined by the latter. In 1972 Gerasimos Santas expounded this fact very clearly (and kindly set the present writer straight about it).<sup>12</sup> More to the present point, that same conjunct would rule out the testing of candidate definitions that are ubiquitous in the socratic dialogues. If one didn't start out with something assertible for the testing, there wouldn't be any way of or point in pretending to engage in the testing in the first place.

The testing of candidate definitions, notably the testing of them that ends in refutation, requires that we start out willing to employ *something* to test the candidates. Vlastos and others respond to this fact by searching for marks or criteria to indicate that a judgment is usable as an elenctic premise. They ask, to repeat, what can distinguish the judgments used as premises in the refutations from the candidates that are thereby defeated? They then look for distinguishing criteria.

Here, however, we must attend carefully to what we're up to. According to the sort of view that Vlastos defends, the marks that we employ must obviously do more than simply be coextensive with the judgments that are accepted as premises in socratic refutations. They must also provide a *justification* for that acceptance. For according to Vlastos' construal of the point of the elenchus, a justification is required for saying that the proposed definition is wrong. That requires that the premises used against it be true. So we seemingly need a justification for that ascription of truth.

The scholarly literature, I'd say, contains no examples of criteria that adequately serve this purpose. I won't pause here over the ones that are on offer. A good example of what can go wrong is provided by Vlastos' own startling suggestion (1994, pp. 24–8), that Socrates holds that his own beliefs are true. Since no one (especially not the self-examining Socrates) would offer "I believe that *p*" itself as a justification of "*p* is true," Vlastos suggests that it's given its status as a justification by an argument whose *assumptions* are that Socrates' beliefs are consistent, and that everyone holding a false belief holds inconsistent beliefs. Since Socrates doesn't try to justify either of *these* claims (and since he often shows how easy it is for an inconsistency to lurk in someone's beliefs without being noticed, even when the domain is restricted to moral or evaluative beliefs), it's impossible to see how he could suppose that this line of thought provides the desiderated criterion.<sup>13</sup>

I don't see any good reason to maintain that the socratic dialogues themselves even attempt to provide a criterion of truth or anything like it at all (cf. Benson, pp. 47–52). Not only do they not present a single proposal for one but also, even if they did, the establishment of a criterion would require a lot of careful discussion of the whole issue—of which these dialogues exhibit absolutely none. We would have to recognize—even if we thought that these works need to justify certain assumptions as premises for refutations of definitions—that they're simply not in the business of showing how to do so. I'll argue in a moment that for the purpose of defining, and of testing definitions, such a criterion isn't what's called for.



## 2.4 “Say What You Believe”

But now we seem to be in a very blind and very dark alley indeed. We need premises for refutations of definitions, since the dialogues *say* that there *are* such refutations, and refutations need premises. On the other hand the dialogues don’t give any criterion for picking out the premises, let alone a criterion for saying which are true. In particular, the dialogues don’t hint at a criterion for picking out, for use as premises, judgments that are true. What are we to do?

The right response, I think, comes from those who attend sufficiently to the one thing that these works explicitly say about what the interlocutors should employ for the scrutiny of a proposed definition. It’s this: “*Say what you believe!*” Vlastos recognizes this fact (1994, pp. 7–8, 14; cf. Benson, p. 53) but doesn’t, I think, exploit its true potential. It amounts to this: “Use the judgments that you actually hold.”

Using the judgments that you actually hold has to mean, of course, use the judgments that you bring to the discussion. The tailoring of one’s answers or of one’s beliefs to preserve consistency with the definition that’s been proposed is explicitly ruled out (e.g., *Rep.* 351e, 339d–340a). Thus, as will become clearer later, the judgments that one actually makes constitute the constraint against which the candidate definition is explicitly to be tested.

We’re looking, remember, for—*inter alia*—a way of characterizing the judgments that may properly be used as a premise against a proposed definition. These, we should conclude, are the judgments that constitute the constraint on an acceptable definition. A definition that conflicts with them is rejected. The important thing to see now is that this injunction to “Say what you believe” provides just the characterization and just the constraint that we need.

But this injunction *doesn’t* provide the constraint by providing a *criterion* of truth. As I’ve already noted, it would be mad to offer “is believed by me” (or even “is believed by Socrates”!) as expressing a criterion of truth. Moreover, as also noted, no criterion of truth is sought or argued for. Therefore we should try to understand a different way in which “Say what you believe!” can provide the constraint that it does.

What, though, could be the rationale for this constraint? Certainly it must have *some* justification in order to form such an important part of the elenctic procedure.

As we saw, Vlastos insists in effect that a constraint on the judgments that serve as elenctic premises refuting failing candidates must consist in their ability to lead us to a definition that’s *true*. There’s nothing to quarrel with in that. It turns out that the Socratic procedure is compatible with requiring, in the end, that an acceptable definition be true in as strong a sense as you like.

However, Vlastos is mistaken about *how* we’re to try to reach a definition that meets this requirement, i.e., is true. For requiring that doesn’t in the least entail that the way in which the premises lead us to a definition that’s true must necessarily consist in *their being themselves true*, or for that matter in showing that each failing candidate is *false*. That this entailment holds Vlastos seems to have taken for granted, e.g. in passing seamlessly from the one thing to the other (e.g., at 1991, p. 114 and 1994, p. 4) without any warning that he’s drawing a substantial inference.

How then can inadequate candidate definitions be disposed of without the need to show them to be false, and without the need to show assumptions used against them to be true?

By taking them, I've suggested, to be ruled out by the constraint expressed in "Say what you believe." What, then, does this constraint amount to? Vlastos asserts that it amounts to requiring "honesty in argument" and "seriousness in the pursuit of truth" (1994, pp. 7–9). However, the constraint amounts to something else too, something significant for the methodology of the *elenchus*.

We can see how this can be so if we consider the purposes that a definition is supposed to serve and the context in which it's sought. Why, we should ask, are definitions sought at all? Not just for intellectual interest or for fun. And even if they're sought for determining, in the end, how things really are, that's not the aim that the dialogues present as proximately precipitating the search for them.

Rather, in the socratic discussions, they're invariably sought in order to provide the *guidance* that's deemed to be necessary for the investigation of various judgments that we need to investigate, which in turn are necessary for decisions about action.<sup>14</sup>

It's continually emphasized in these works that we need definitions to investigate certain issues. Moreover all of those issues have something in common. They're all issues that are *problematic*. All of these works start with a practical issue to be settled. All of these issues turn on making a judgment that's subject to *uncertainty* or *disagreement*. That's why a definition is called upon, and indeed said to be indispensable.<sup>15</sup>

Illustrations of the need for definitions in order to provide such guidance are as easy to find as definition-seeking socratic dialogues. For instance: we need to define courage (in the *Laches*) in order to gain guidance about whether a certain kind of training will instill courage; we need to define justice (in *Republic I*) in order to tell whether being just will benefit us, and so to decide whether to undertake to be just; we need to define piety (in the *Euthyphro*) so that we may determine whether a particular action or kind of action (e.g., Euthyphro's prosecution of his father) is pious, and so to undertake it; and so on.

It's no surprise that definitions are sought for the sake of guidance in making and accepting judgments. It's fully explicit in these dialogues that the incapacity to perform the judgment-guiding function disqualifies a definition. One sort of problem for a definition arises, for example, when it's uninformative because it's circular (*Meno* 79b–d). Another arises when a definition is uninformative in another way, by virtue of containing a term that's not understood (*Meno* 75b–c). Yet a third reason for rejecting a definition is its leading to mutually conflicting judgments involving the target term (*Euthyphro* 7a–8e). No contrast is drawn between this kind of refutation and a refutation by, for instance, counterexamples. In both cases the definitions are to be rejected. (For this reason "refutation" in a logical sense isn't the right translation of the Greek word "*elenchos*.")

We here see, by the way, a difference between an elenchus directed against a proposed definition and an elenchus used to refute something of a different kind (e.g., some of those mentioned by Vlastos 1994, pp. 11–2). Except for such

non-definitional theses as are designed to help directly in constructing a definition, these other targets of elenchus don't have the guiding function that definitions do.<sup>16</sup>

When we encounter disagreement and uncertainty about problematic judgments, and wish to gain guidance in settling our minds about them, where do we begin? For this we need to establish a definition, not just produce one. How is that to be done? A plethora of different definitions are possible, having different consequences for the problematic judgments.

Consider this. Although some judgments are indeed problematic, people do in addition hold many further judgments about which doubts and disagreements haven't come to light. That has to be so, just as Geach insisted; and it is so in these dialogues, as Santas showed. If it weren't so, then when Socrates enjoins, "Say what you believe," people simply wouldn't have anything to say.

As a matter of everyday fact, people do make judgments, and they do so before they ever even consider searching for definitions. As Socrates' interlocutors are pictured, they arrive at the conversation prepared to advance plenty of judgments involving terms that are to figure as *definienda*. On the other hand they have trouble passing verdicts on problematic judgments. About some issues they're uncertain or they disagree—e.g., how to become courageous or virtuous.

On the other hand, most of the interlocutors aren't initially familiar with the business of formulating definitions. Sometimes Socrates has to explain to them what it's all about, and presses them into it despite their resistance. People normally, after all, use terms without definitions.

As the dialogues are actually constructed, the terms that they treat aren't taken to be completely incomprehensible to the interlocutors. The interlocutors regard themselves as capable of using the terms that become targets of investigation. This capacity is exhibited in the interlocutors' readiness to make certain judgments involving these terms. It's hoped that definition will help us move from this condition to the making and evaluating of further, *problematic* judgments.

As a philosophical matter we can describe the situation this way. The interlocutors begin with what they think is some grasp of the term at issue, as articulated in judgments that they bring to the discussion, but they come to recognize that their grasp of it isn't sufficient to allow them to assess all of the judgments in which the term figures (see e.g., *La.* 194a). In particular, they can't confidently and unanimously assess the problematic judgments.

Thus the sought-for definition can be thought of as a means—the only one that the socratic dialogues consider—for *extending*, so to speak, one's understanding of the term so as to cover the problematic judgments by determining how they are to be evaluated. Before we have the definition of courage, in a certain sense we don't understand courage fully enough to determine which kind of training (if any) will produce it; once we've established the definition, we suppose, we'll be able to determine that. So we start with what we can call in *that* sense an *incomplete* or *partial* grasp of the term, which we then want to extend.<sup>17</sup>

This gives us a sense of why a conflict between a proposed definition and judgments used to refute it generates a refutation of the former and not the latter. In the minds of the interlocutors, the judgments are there first. Then the definitions are

sought in order to build out to further judgments. To put it another way, the definition is required to extend what's already been partially grasped and articulated by the judgments that are brought to the discussion.

Some of Plato's readers have thought that the definition is required not merely to extend one's understanding of the term, so as to guide us in deploying it in problematic judgments, but also to justify and defend its deployment in *all* judgments, including the ones in which it's been used heretofore. *No* judgment can be advanced or used, on this view, until a definition is established.

Here's Geach's "Socratic Fallacy" revisited. If things were like this, then as Santas showed we'd have nothing to use to test the definition in the first place—failing, that is, some operation to bootstrap ourselves up from judgments that aren't adequately supported when we enter upon the discussion, to judgments that can then be accepted after definitions are in place. I don't think that the dialogues that I'm discussing go in for any such bootstrapping, or even foreshadow it.<sup>18</sup> As I've said, the judgments that one brings to the discussion, which are not to be tailored to conform to the definition, constitute the constraint against which it's to be tested. It's hard to see any sense at all in that injunction to "Say what you believe" if the interlocutor's judgments aren't to play a constraining role on an eventual definition.

## 2.5 Elenchus Is Only the First Test

Suppose that the foregoing account is accepted. Before turning to a line of objection against it, I want to consider its implications for what it would take to *refute* a definition. These implications give us a clear view of what definitional elenchus should reasonably consist in, but without the contortions required by the kind of interpretation that Vlastos put forward.

First notice: nothing said so far depends on regarding the judgments that one brings to the search for a definition as *true*. They're *simply* the judgments that the parties to the discussion *have* at the start, which involve the target term, and that in *that* sense (and that sense alone) articulate the partial grasp that one has of it.

Now consider the implications of discovering that one's proposed definition of a term conflicts with judgments involving the term that one accepts. That discovery would mean that the aggregate that one has at one's disposal to evaluate problematic judgments involving the term is inconsistent.

If that aggregate is inconsistent, however, then, as it is, it can't perform its guiding function. It won't help you, as you survey the range of judgments involving the term, determine which are correct and which aren't. If it won't do that, then it isn't satisfactory as a guide, and therefore can't serve the purpose that's its *raison d'être*. Thus a definition that conflicts with accepted judgments involving the term that it's supposed to define can't be acceptable as a definition.

That means that the aforementioned inconsistency is *all* that's needed in order to show that a candidate is to be rejected as a definition. *Nothing further* is required. In particular, it's not required that one demonstrate the truth of the assumptions that are shown inconsistent with the definition.

If this is the way in which the scrutiny of socratic definitions is supposed to proceed, then *no wonder* that in these dialogues the premises used against candidate definitions aren't proved, and that no criterion is given (or even sought or mentioned) that might establish their truth.

Not, of course, that an elenchus is unconnected with truth. Indeed, as I'll urge in a moment, the elenchus can be linked to truth construed as strongly (and "realistically," for that matter) as you like. Moreover inconsistency itself is a matter of being incapable (whether syntactically or semantically—not that Plato ever had the distinction) of being *true together*.

The line of objection that must be considered, however, says that *this* particular kind of connection with truth isn't enough. Following the spirit of Vlastos' account, the objection contends that unless the scrutiny of a definition directly involves saying that the premises used against it are true, it won't be leading us to a definition that's true in the required way. At best, the procedure would in the end show no more than mere conformity to the beliefs of the interlocutors, or of their speech community or its conventions, or something of that kind. That might be all right for mere "nominal" definitions, it will be said, but not for the definitions that Socrates is after (even if he doesn't use Plato's doctrine of Forms to say so).

This line of objection is mistaken. In the first place, it mistakes a sufficient condition for the correctness of a definition with a sufficient condition for rejecting an inadequate candidate. As a result, the objection ignores the question what further conditions might turn out to be necessary for the correctness of a candidate definition, over and above its surviving the elenctic tests that the socratic dialogues exhibit.

The difficulty with the objection can be put like this. When a proposed definition can't serve as a guide to judgment because it's inconsistent with the judgments used to test it, it can't even pass the *first test* for adequacy. That's *all* that an elenchus needs to show in order to dispose of the proposal. And—just as Vlastos said—it's all that an elenchus *does* do. To convict the candidate of inadequacy, we don't need to display *other* virtues that it doesn't have or requirements that it doesn't fulfill.

Saying that, however, is fully compatible with going on to say whatever one wishes about the *full* requirement for a definition to be correct. Vlastos and others can be as emphatically correct as you like about whatever other such requirements—of correspondence with reality or Forms, or infallibility or incorrigibility—you might choose. Which of these Plato wishes to impose, in these dialogues or subsequently, is of course a vexed question.

Bear in mind that there seemingly *must* be something more to the establishing of a definition as true than its merely passing an elenctic test, or being consistent merely with the judgments of an interlocutor that one happens upon. By common agreement, an individual elenchus isn't like a proof procedure; it's like a *disproof* procedure. The same is true of any arbitrary set of elenchi. Whether Plato ever envisages some set or totality of elenchi, the passing of which would establish a definition (as one might perhaps infer from *Rep.* 534b–c), is unclear. But I think it's safe to say that no such thing plays any role in the socratic dialogues.

Thus within the sphere of definitions elenchus is a scrutiny of a proposal that can refute it but isn't claimed to be able to establish it. That's why it makes sense to suppose that to flunk an elenchus is, as I put it, to flunk the *first* test, or in other words that a *full*, sufficient condition for a definition to be correct needn't by any means be the very same condition an elenchus shows it not to fulfill.<sup>19</sup>

## 2.6 Specifying the Definiendum

Nevertheless, the line of objection hasn't yet come to an end. It's one thing to explain the compatibility of taking the ultimate aim of elenchus to be a true definition and supposing that, nevertheless, the elenctic premises needn't be shown to be true. That's what I've just done. It's another thing, however, to explain how the mere fact of a judgment's having been brought to the discussion makes it usable for disposing of a proposed definition. This latter task remains to be undertaken.

Interlocutors come to the discussion with judgments—containing the relevant term—that they accept. The attempt to find a definition that allows us to extend the deployment of that term comes later. This definition is supposed at a minimum to fit—not conflict with—the prior judgments.

The objection points out that, after all, being chronologically earlier isn't much of a status. Why should *that* determine what's refuted by an elenchus? Perhaps we should hold onto the candidate definition, and reject a premise that's used against it.<sup>20</sup> Why not?

The original rationale of the candidate was precisely: to extend *that* term, which has been partially articulated by the judgments that have been made with it. Such credibility as the candidate had was as an attempt to do that. Nothing gave it any status, as compared with those judgments, that would give one reason to jettison them in favor of it. It didn't succeed in doing that, as the ensuing contradiction revealed. So unless the candidate is provided with some independent source of credibility, there's no choice but to reject it as a definition of *that* term.<sup>21</sup>

But why take *those* judgments—the ones that it seems one just *happens* to bring to the socratic conversation—as the articulation of the term to be defined? It's of course not an option to say that in place of those judgments the candidate definition should be accepted. For the falsity of those judgments wouldn't imply the truth of the candidate (they're not contradictories, since both of them might be false; cf. n. 6).

The reason why *those* judgments, namely the ones that are in fact accepted, are taken to articulate the term to be defined, and thus to constrain the definition, is simply this: they're *all that one has to go on* in determining what to search for. What other judgments should one take? Judgments that one *doesn't* accept?

We have to take *something* to determine what's to be defined. Which term do we want to define? "Courage," let's say. What term is *that*? Well. . .—and here we have nothing to say *except* what we can give by adducing judgments that we make involving it. Without them one has nothing which the definiendum would consist in.

As Peirce put it, “We begin where we are.” Another option might *seem* to be to say, at that juncture, that *we just don’t know what* term we’re trying to define. But that removes the possibility of a specifiable search altogether. So the only open option is to regard our judgments as doing the specifying.<sup>22</sup>

These options, by the way, are the ones that the *Meno* alludes to in presenting its paradox at 80d–e.<sup>23</sup> How that dialogue and subsequent works think to respond to the paradox is another vexed question, but again not a question that the socratic works address.

The claim being made here is that the judgments that the interlocutors accept *have* to be regarded as constraints on any proposed definition if there’s to be a search for one. The claim *isn’t* that the elenctic judgments articulate the definiendum term *truly*. Rather it’s that *if* one has available to oneself *anything* that can lead one to an articulation that extends to the problematic judgments, it’s these judgments or some of them. Whatever chance one has of defining correctly depends on making what one can of the indications that such judgments supply.<sup>24</sup> A skepticism directed toward them as such would therefore make the search for a definition impossible from the start, just as Geach saw (and Plato too, at *Meno* 80d–e).<sup>25</sup>

It’s certain at all events that whatever definition one eventually accepts will have to be compatible with whatever judgments involving the definiendum one *then* makes at the time of accepting it. But the judgments that one accepts *now*, in the midst of a search for the definition, are all that one has to indicate what one’s eventual judgments at that juncture might turn out to be. Once again, as Peirce says, we begin where we are. That’s because the *whole question* (even if one takes it, as Socrates didn’t, to be a quite general methodological question) is about how, starting as we are, we might rationally conclude a search for a definition successfully.

This isn’t to deny that in Plato’s view (and perhaps even in philosophical fact), the acceptability of one’s judgments has at some point to be *explained* somehow by their connection to some “reality” that’s “independent” of them. Indeed, subsequently Plato by common agreement wished to defend such a claim. He may even have foreshadowed it here—who knows?<sup>26</sup> And Vlastos may even be right to seize on a passage in the *Gorgias* that could be taken to insist on it, or on some status close to it.<sup>27</sup> But that’s not in the socratic works.

## 2.7 The Group

One rationale for regarding consistency with the judgments that a person accepts as a threshold test for a definition can be seen if one considers that often noticed fact that in the socratic dialogues, the elenchus is partly a *personal* test—that is, in this kind of case, a test of the *person* who’s trying to formulate an adequate definition. We should think about why that should be.

What’s the upshot of discovering, through the wealth of illustrations that these works supply, that people routinely fail to offer definitions of important terms that are consistent with their actual judgments? Consider what kinds of people are covered by this discovery. They form a broad cross-section of Athenians. They



include, most especially, ordinary people with no special claim to expertise and also experts in various areas, including such eminent figures as Protagoras and Gorgias. They also include both iconoclasts like Callicles and Thrasymachus, and also very conventional people as well, like the gentlemen in the *Laches*.

Showing this broad group of people incapable of consistently capturing the terms that are partially articulated by their very own judgments would be of enormous philosophical significance. It would mean that the resources for settling problematic issues can't be directly culled from the judgments either of ordinary speakers of the language or those of experts, or both together.<sup>28</sup> That seems clearly to be an important point that the socratic dialogues illustrate, and are intended to illustrate.

It should be mentioned briefly at this point—though it deserves extensive discussion—that the scrutiny that takes place in a socratic work isn't simply the scrutiny of a single individual or interlocutor. Instead it's the scrutiny of a *public group*, which normally includes Socrates himself (see e.g., *La.* 200e–201a; *Prt.* 333c: “. . . it may turn out that *both the questioner, myself, and my interlocutor* end up being tested”). We're not dealing here with a private language, nor with an individualist conception of understanding or of elenctic scrutiny (see again *Rep.* 534b–c).<sup>29</sup>

But since the task of consistently articulating and extending the deployment of a problematic term outruns the capacities of a public of this kind, including the judgments of both ordinary speakers and assorted (*soi-disant*) experts, one should conclude that if there's any way of settling the problematic issues that definitions are intended to guide us through, it must be a much more thoroughgoing way than what the elenchus' culling of such judgments can offer. Surely it makes sense to suppose that this was Plato's view—at least retrospectively, when he wrote his subsequent works introducing and exploring the doctrine of Forms, and quite possibly already when he wrote the socratic definition-seeking dialogues as well. That hypothesis ascribes to the elenchus an appropriate role, free from confusion, in Plato's perception of what might be required for understanding.<sup>30</sup>

## Notes

1. The two pertinent works by Vlastos are his 1991 (a reworking of an earlier paper, which appeared in 1983) and his 1994.
2. I use the lower case “socratic” because—even when strong convention imposes the upper case “Socrates”—I'm not concerned at any point with the historical Socrates, but only with what goes on in his name within certain Platonic works.
3. Thus the works that fall directly within my purview here are the *Laches*, *Charmides*, *Lysis*, *Hippias Major*, *Euthyphro*, *Meno* up to 80d, and *Rep.* I. This choice of dialogues is intended as thematic, not as necessarily chronological. The set of “socratic definition-seeking dialogues” here is united by the presence of a search for a definition, the failure of that search through the refutation of candidate definitions, and a statement of the unfortunate consequences of the failure. The first part of the *Meno* fits this specification: it's structurally the same as the other socratic dialogues until 80d–e, at which point it generalizes certain themes, not present in them, about how a successful search for a definition might or might not be possible. Like many scholars, I take *Rep.* I to be a retrospectively produced sample of a socratic definitional



dialogue, which sets the stage for a differently conceived approach in the rest of the *Republic*. By theme and systematic import the *Lysis* obviously counts as a socratic definition-seeking dialogue. Vlastos' exclusion of it—Vlastos (1994, pp. 30–1; 1991, p. 128)—is based on considerations irrelevant to those that concern me. Other dialogues (seemingly from the same roughly defined chronological period) direct our attention to the same systematic need to find definitions, though as is widely recognized they range farther afield and advance more substantial positive views. These include the *Protagoras* (361a, d) and the *Gorgias* (of rhetoric, 447c, 448e). Vlastos' inclusion of the *Gorgias* with the “elenctic” works is based, again, on irrelevant criteria; it also has the effect of inducing him to use an idea that occurs in no definition-seeking dialogue as controlling for his interpretation of elenchus, which I think distorts the picture; cf. Benson (2000, p. 53 n. 72, pp. 80–5).

4. Most of the elenchi directed at non-definitions seem to me defensible for reasons analogous to the ones adduced *infra* concerning definitions. Many of them are directed explicitly at elements of proposed definitions (e.g., *Laches* 198a with 199e). In the other cases, a sound refutation isn't in fact produced (we shouldn't rule out that sometimes Socrates has simply not delivered what he needs to).
5. For the most part Vlastos claims that the socratic discussions deal only with “moral” (I'd prefer to say more broadly “evaluative”) matters (1994, pp. 6–7, 26). I don't think that that claim is accurate, and I think it's philosophically important that such a distinction isn't drawn here (*Euthyphro* 7b–d seems to me to be about a quite different distinction). But I'll have to leave this issue aside.
6. One could get the impression from Vlastos' exposition that a candidate definition and the assumptions used against it are mutual *contradictories*. That can't be right. If it were, then the option of denying the latter would require one to accept the former. It's obvious, however, that in the normal case the candidate and the conjunction of assumptions, though they couldn't both be true, could both be false (since a different definition might be true). See *infra* note 20.
7. I prefer “judgment” to other comparable expressions, partly because it's unencumbered by as much surrounding epistemological or semantic theory (unlike, e.g., “belief,” “proposition,” and “statement”).
8. This point is particularly important. It leads some interpreters to try to uncover a method that Plato can be taken to attribute to Socrates (either the character or the historical personage). In my opinion, that's a mistake; rather, Plato intends to portray methodologically unself-conscious discussions, mainly to display the kind of outcome that they can be expected to have.
9. I think that Santas (1979, p. 84) and Benson (2000, pp. 3–6, 189–90) are right to oppose these interpreters (who include Vlastos 1994, p. 3).
10. For treatments of these issues see Benson (2000, *chaps. 6, 7*). For reasons given in part here, I don't accept Benson's view that a principle of the Priority of Definition is at work here; but I can't here take up the whole range of arguments for it.
11. Notice that this construal of the situation overlooks doubts that might attach to the legitimacy of inference—licenses themselves, and attends only to premises and proposed definitions.
12. It looks to me that Santas modified his view somewhat in his (1979, pp. 311–2), though I tend to think that his position in his (1972) was right.
13. For some criticisms of Vlastos' line of thought see Benson, pp. 37 ff. As to another proposal: I don't think that it helps here to read the distinction drawn in the *Meno* between knowledge and true belief back into the socratic definition-seeking dialogues, since they don't mention or (I think) even hint at it or show it to be at stake (see also Gareth Matthews 1999, pp. 43–5). Benson is right that at times in the socratic works a distinction is presupposed between knowledge and true belief (pp. 93–5). It does not seem to me, however, that the distinction is *used* for the purposes relevant here. For, as a philosophical matter, the distinction plainly won't help with Vlastos' problem or Geach's. Like the socratic works, they're concerned with the question how we could identify a judgment as “true” *in advance of* having established a definition. Saying that at *that* juncture we don't need knowledge but only true belief simply lands us in Vlastos' problem all over again—or rather, fails to get us out of it. (It's another question

- whether saying *ex post* that we *had* true judgments would form part of a helpful explanation of how we could reach a definition. Unfortunately there's no space to explore this matter further here.)
14. Surprisingly Richard Robinson's *Plato's Earlier Dialectic*, for instance, which so helpfully initiated the current discussion of all of these matters, neglects (so far as I can discover) to broach the question why definitions are sought—so fixed a place defining has come to have as something that philosophers simply *do*.
  15. I'm prevented by lack of space from treating a number of issues often addressed in discussing the socratic searches for definitions. These include: (1) whether any knowledge involving a term can be claimed prior to establishing a definition of it; and (2) whether establishing a definition is *sufficient* for determining the truth of any judgment involving it.
  16. One of the causes of the false leads generated by Vlastos' type of account is a failure to attend sufficiently to the application of elenchus to definition (cf. n. 4).
  17. The notion of "incomplete grasp" is used by Burge in his treatment of ideas of Frege that are importantly similar to those of Plato. The deep affinities between the two thinkers deserve extensive study (for relevant points in Frege, see the index references for the term "understanding, incomplete," at Burge 2005, pp. 418–9, and compare pp. 252, 262 with, e.g., *Rep.* 432b–e, which has more significance than it has been credited with). I take it that in Plato, as opposed to Frege on this point, one's grasp of the term includes, so to speak, the *taking* of one's grasp *to be* incomplete, insofar as one takes there to be judgments involving it that one can't adequately evaluate but in some sense should be able to. But this is a complex issue.
  18. Various operations have been proposed. Some of them are "coherentist" (e.g., Benson 2000, ch. 9; and perhaps Irwin 1977, pp. 69–70, 140). Others have a different character (e.g., seemingly, Irwin 1995, pp. 126–36). If Plato favors such an operation, the *Meno* after 80d is the first place where he does so. The socratic dialogues and the first part of the *Meno* describe the socratic definition-seeking project *before* any such strategy (if it's genuinely there at all) has been applied. One important reason for making this separation is the following. Even in the *Republic* one's taken to be able legitimately to reject a rival false account of what something is before the correct account is reached (see e.g., *Rep.* 505b–506d with 534b–c, and the *Tht.* as a whole). Even after the socratic works, Plato seems to believe that definitions and other such identifications can be shown false even before the correct one has been shown true (whether that latter be done by an argument involving coherence or not).
  19. Here's another indication that, as noted earlier (n. 4), there's a large difference between an elenchus that's directed against a proposed definition and one that's directed against another sort of thesis. In the latter case, the elenchus is typically the end of the matter; in the former case—if one really hopes to find a correct definition—it can't be.
  20. As mentioned in n. 6, the candidate definition and the premises could both be false (since a different definition might be correct), so we can't infer the definition's truth by rejecting (some of) the premises.
  21. The idea that such an independent source of credibility is possible is broached by the passages introducing the "method of hypothesis" in the *Meno* and *Phaedo*: the derivation of a hypothesis from a "higher" hypothesis. It's also broached by Callicles in the *Gorgias* and Thrasymachus in *Rep.* I, because they both have *theories* about how the ordinary usage of the terms "moderation" and "justice" has come about, which gives their definitions a claim to independent credibility that none of the other candidate definitions have. It's no accident that Thrasymachus' opposition (including his denial that justice is a virtue at all) is what precipitates the entirely non-socratic attempt to explain justice that takes up the following books of the *Republic*. Nothing like this happens in the works under examination here.
  22. *All of them?* Does this mean that every judgment involving a term plays an essential role in specifying it, so that none of them can be revised? No. But at the present stage of a socratic discussion, we have no basis for distinguishing between those that are dispensable and those that aren't. After further discussion, that could change, as it does for example in *Rep.* V, where Glaucon and Adeimantus revise their view that it would be wrong to hold wives and children

- in common—this being clear evidence that now a definition is held to have independent credibility (cf. previous note).
23. The former, the paradox at *Meno* 80d–e suggests, entails saying that we already “know” the definiendum. The solution consists in explaining why this entailment, if accepted, isn’t damaging to the project of defining.
  24. Once again, *Rep.* 435b–e helps somewhat to see how Plato views the matter.
  25. It’s probably a good idea to stress that the problem here at bottom *isn’t* an *epistemological* one, about how one *can tell* what term one’s trying to define. It’s a problem about how there *can be such a thing as* the term that one’s trying to define. The issue is analogous in this way to the one clarified by (Stroud 2000, pp. 75–6; see also his reference to Goldberg).
  26. This is the view of, just for example, Reginald Allen. One reason supporting it is the fact that in the *Euthyphro* itself (6d), Plato says that pious things are “made to be” pious by piety. That suggests an *explanatory* capacity, which arguably might have to be in some relevant sense independent of judgments involving it.
  27. See esp. *Grg.* 479e, 508e–509a. Even if Vlastos is right about the meaning of these passages, there’s no good reason to say that it is part of the thinking presented in the socratic definition-seeking dialogues themselves. It’s plain that the *Gorgias* brings more ideas into play than what those works deploy, and uses them in a different way (cf. again Benson 2000, pp. 80–5).
  28. This is the point that Burge makes about Frege’s views on mathematics (esp. p. 261: “. . . cases where the most competent speakers, and indeed the community taken collectively, could not, even on extended ordinary reflection, articulate the ‘standard senses’ of terms”; and p. 276: “Nothing in the current mathematical usage determined how to extend the conventional ‘meaning’ of ‘Number’ to cover cases about which it was previously mum. Further philosophical and mathematical considerations had to be appealed to”).
  29. This point is, I think, relevant to the issue raised by Benson (2000, pp. 37–40).
  30. Many thanks to Hugh Benson and Gareth Matthews for comments on an earlier draft.

## Bibliography

- Allen, R.E. 1970. *Plato’s Euthyphro and the Earlier Theory of Forms*. London: Routledge.
- Benson, H.H. 2000. *Socratic Wisdom*. Oxford: Oxford University Press.
- Burge, T. 2005. *Truth, Thought, Reason: Essays on Frege*. Oxford: Clarendon Press.
- Irwin, T. 1977. *Plato’s Moral Theory*. Oxford: Clarendon Press.
- Irwin, T. 1995. *Plato’s Ethics*. Oxford: Oxford University Press.
- Matthews, G.B. 1999. *Socratic Perplexity and the Nature of Philosophy*. Oxford: Oxford University Press.
- Robinson, R. 1953. *Plato’s Earlier Dialectic* (2nd ed.). Oxford: Clarendon Press.
- Santas, G.X. 1972. “The Socratic Fallacy.” *Journal of the History of Philosophy*, 10:127–141.
- Santas, G.X. 1979. *Socrates: Philosophy in Plato’s Early Dialogues*. London: Routledge.
- Stroud, B. 2000. “Wittgenstein’s ‘Treatment’ of the Quest for ‘a Language Which Describes My Inner Experiences and Which Only I Myself Can Understand’.” In edited by *Meaning, Understanding, and Practice*, B. Stroud. Oxford: Oxford University Press.
- Vlastos, G. 1983. “The Socratic Elenchus.” *Oxford Studies in Ancient Philosophy* I:27–58.
- Vlastos, G. 1991. *Socrates Ironist and Moral Philosopher*. Ithaca: Cornell University Press.
- Vlastos, G. 1994. *Socratic Studies*. Cambridge: Cambridge University Press.

# Chapter 3

## Reasons and the Problem of the Socratic Elenchos

Alejandro Santana

### 3.1 Introduction

In this essay, I address what Vlastos called *the* problem of the Socratic elenchos. Vlastos stated the main question of the problem as follows: “how is it that Socrates claims to have proved a thesis is false when, in point of logic, all he has proved is that the thesis is inconsistent with the conjunction of agreed-upon premises for which *no reason* has been given in that argument?”<sup>1</sup> This problem has long been important to solve, but it has been the subject of a good deal of controversy: not only is there controversy about how to solve the problem but there is also controversy about whether there is a genuine problem in the first place.<sup>2</sup>

The purpose of this paper is to address the *latter* controversy, part of which centers around an important but ambiguous assumption which seems to require that Socrates gives or has *particular* epistemic reasons for the truth of the premises he uses to refute the interlocutor. It is important to address the controversy regarding this epistemic assumption. The problem is generated by *both* this assumption *and* the constructivist assumption that Socrates thinks his elenctic method can establish the falsehood of the interlocutor’s initial claim.<sup>3</sup> If this epistemic assumption is unjustified, then the problem is, at least in part, groundless.

This assumption, however, has not received as much attention as the constructivist assumption, though it is no less important. More specifically, little work has been done to make explicit why exactly this assumption is made, how it should be understood, and whether it is justified. More importantly, little work has been done to see whether these considerations have any implications for an adequate solution to the problem.

I plan to address this part of the controversy and, in the process, do this other important work. To do so, I will first present the problem and its epistemic assumption.

---

A. Santana (✉)  
University of Portland, Portland, OR, USA  
e-mail: santana@up.edu

Second, I will present how this assumption has been interpreted and show how these interpretations ground serious objections, raised by Kraut and Santas, to the legitimacy and scope of the problem. Third, I will argue that the epistemic assumption avoids these objections and, when properly understood, is justified. I will thereby argue that the problem is well-grounded, at least with respect to *this* assumption.

To argue this, it is important to briefly revisit the kind of problem that Vlastos had in mind when he raised it. I think that doing so in light of these two objections will help advance the scholarship on this problem in the following ways. To begin with, it will help clarify the epistemic assumption upon which the problem of the elenchos is grounded. I aim to show that Vlastos raised a general methodological problem about the epistemic status of the auxiliary premises. Given this, Vlastos expected Socrates to give or have general epistemic reasons for the premises he uses, not particular ones. Such reasons could serve to ground the particular auxiliary premises he uses, but they would serve primarily to ground Socrates' elenctic method by revealing how the auxiliary premises have a weightier epistemic status than the interlocutor's initial claim.

Revisiting the way Vlastos raised the problem will also shed light on a facet of the problem that, to my knowledge, has not been fully appreciated in the scholarship. I aim to show that, while Vlastos may have overstated the point that Socrates gives "no reason" for the auxiliary premises, he was right to point out that Socrates elenctic method depends on the interlocutor's sincere agreement to secure them, which Socrates makes clear in the *Gorgias*. Moreover, Vlastos was right to point out that this raises a problem for the elenchos because in the *Crito* (44c7, 48a8) and *Gorgias* (471e2–8, 474a2–9) Socrates dismisses the agreement of others as having no credibility. Consequently, Socrates' methodological dependence on the interlocutor's agreement seems to be *inconsistent* with his own expressed dismissal of the agreement of others as having no epistemological weight. In both cases, it's *mere* agreement. Given this apparent inconsistency, it is quite legitimate to ask Socrates how he secures the auxiliary premises in his elenctic arguments.

What is more, this facet of the problem will provide insight on a novel strategy to adequately solve the problem of the elenchos. In the final part of this essay, I will offer a sketch of this strategy, provide evidence for its plausibility, and outline a few of its explanatory advantages.

### 3.2 The Problem and the Epistemic Assumption It Makes

The problem, briefly explained, is this. According to Vlastos (1994a, p. 11), a "standard" elenchos goes as follows<sup>4</sup>:

- (1) The interlocutor asserts a thesis, *p*, which Socrates considers false and targets for refutation.<sup>5</sup>

- (2) Socrates secures agreement to further premises, say  $q$  and  $r$  (each of which may stand for a conjunct of propositions). The agreement is *ad hoc*: Socrates argues from  $\{q, r\}$ , not to them.
- (3) Socrates then argues, and the interlocutor agrees, that  $q \& r$  entail not- $p$ .
- (3a) Socrates thereby proves that the conjunction of  $\{p, q, r\}$  is false; either  $p$  is false, or  $q$  is false, or  $r$  is false.
- (4) Socrates then claims that he has shown that not- $p$  is true,  $p$  false.

In step (4), Socrates claims to have *proved* his conclusion. By this, Vlastos meant that Socrates thinks he has given the interlocutor “good reason” to think that  $p$  is false, not- $p$  is true (1994b, p. 46). However, in step (2), Socrates only secures the interlocutor’s *agreement* to the auxiliary premises. Vlastos explains, “...the premises  $\{q, r\}$  from which Socrates deduces the negation of the opponent’s thesis in any given argument are *logically unsecured* within that argument. He asks the interlocutor if he agrees, and if he gets agreement he goes on from there” (1994a, p. 13; my italics). Vlastos later says that these premises are “unargued-for” and that “no reason has been given in that argument” (1994a, pp. 18, 21).

The combination of steps (2) and (4) raises the *formal* component of the problem: because of step (2) Socrates cannot validly conclude (4) from (3a). Although the interlocutor agrees to the auxiliary premises, he *also* thinks his initial claim ( $p$ ) is true; hence, from the interlocutor’s point of view, *all* the premises have the same truth-value—they’re *all* true. The problem is that since Socrates secures only the interlocutor’s agreement to the auxiliary premises, he only establishes that the interlocutor thinks they’re true. Socrates, therefore, fails to distinguish the initial claim from the auxiliary premises and thereby cannot justify targeting  $p$  as the false premise. Consequently, Socrates leaves open the possibility for the interlocutor to target *any* of the other premises as false instead of targeting  $p$ . Although the interlocutor usually targets and abandons  $p$  as the false premise, he does not have to: he can retain  $p$  and reject  $q, r$ , or both.

The *formal* component of the problem raises the *epistemological* component of the problem: because Socrates secures only the interlocutor’s agreement to the auxiliary premises, their epistemic status remains *undisclosed*. Socrates and the interlocutor therefore cannot validly conclude (4) from (3a) because it is unclear how the auxiliary premises have a weightier epistemic status than the initial claim does not have.<sup>6</sup> Both Socrates and the interlocutor think that the auxiliary premises are true; however, in order for Socrates to think that they can target the initial claim as false, he must somehow think that the auxiliary premises are *epistemically weightier* than the initial claim. This is especially true for the interlocutor, who must be *talked out* of thinking that his initial claim is true.

This problem is particularly compelling in light of the fact that both Socrates and the interlocutor usually conclude (4) from (3a); hence, they *both* must somehow think the auxiliary premises are more epistemically weighty and thereby think they are *justified* in thinking the premises are true. The epistemological component of the problem, therefore, raises the *central question* of the problem of the Socratic

elenchos: how does Socrates think he and the interlocutor are *justified* in thinking that the auxiliary premises in his elenctic arguments are true?

In light of step (2) and Vlastos' explanation of it, the claim that there is a problem makes the following assumption.

- EA Socrates must give or have epistemic reasons for how he and the interlocutor are justified in thinking that the auxiliary premises in his elenctic arguments are true.

### 3.3 Objections Raised Against the Epistemic Assumption

As mentioned above, this problem has raised substantial criticism, and some scholars deny that there is a genuine problem in the first place. Of these scholars, some have denied the problem because it assumes EA, and they deny EA. This assumption, however, is ambiguous and has been interpreted in one of the following two ways.

- EA1 Socrates must give or have *particular* epistemic reasons for the truth of *every* auxiliary premise he uses in his elenctic arguments.
- EA2 Socrates must give or have *particular* epistemic reasons *not* for *every* auxiliary premise but for *some* of the premises, namely, those that need such reasons.

Some scholars have taken interpretation EA1 due to Vlastos' claim that Socrates secures only the interlocutor's agreement to the auxiliary premises and are thereby "logically unsecured" in the argument. Vlastos also claimed (1994a, pp. 13, 18, 21) that they are "unargued-for," and that Socrates gives "no reason" for the premises in the argument. On this interpretation, Kraut raised the following objection. To give a proof, it is not necessary for one to give or have reasons for the auxiliary premises; one can prove a proposition from an argument containing unsupported premises. Kraut uses Aristotle's infinite regress argument to make his point (Kraut, 1983, p. 62):

As Aristotle realized, if arguments contain a finite number of steps and if circularity is to be avoided, then every demonstration will contain statements for which no argument is given [*An. Post* I 72b5–25]. And if the unsupported premises of one argument are derived from further premises in some other argument, then that other argument will itself contain undemonstrated premises. If the only way to prove a proposition is to deduce it from others, then at some point or other demonstration will have to rest on unproved premises, and it would be unreasonable to criticize an alleged proof on the ground that it rests on undemonstrated assumptions.

If Kraut is correct, then Socrates can prove his conclusion with an argument containing unsupported premises. If so, the problem would be groundless because it assumes EA1, which incorrectly presupposes a stronger concept of establishment than Socrates needs—one that unreasonably requires Socrates to give



or have particular reasons for the auxiliary premises when no such reasons are necessary.

Additionally, Kraut was aware of the central question that the problem asks, and he answered it by arguing that Socrates often uses premises that are so plausible that we can consider them “compelling” and *as yet* in no need of justification (1983, pp. 62, 65). For Kraut, if we want to see whether Socrates gives a proof, we need to check the details of the argument: if Socrates uses reasonable premises, and the premises imply the conclusion, then Socrates has given a proof (1983, p. 64).

On the other hand, some scholars have taken interpretation EA2. Scholars have taken this interpretation in light of several considerations. To begin with, they agree with Kraut that it would be unreasonable to expect Socrates to give reasons for every premise he uses. Yet they also agree with Vlastos’ contention that it would be equally unreasonable if Socrates were *never* to give reasons for the premises. This is especially significant in arguments where Socrates provides no reasons for controversial premises, and there are several such cases.<sup>7</sup> If so, then Socrates would be open to the charge of dogmatism, for it would imply that he assumes that such premises need no reasons. In light of these considerations, scholars have suggested that a more reasonable expectation is for Socrates to give or have *particular* epistemic reasons for *some* of the premises, namely those that need such reasons.

On this interpretation, Santas raises the objection that Socrates sometimes meets this reasonable expectation.<sup>8</sup> In cases where he thinks he establishes that the interlocutor’s initial claim is false, Socrates takes some premises for granted and argues for others. Premises he takes for granted are supported by experience or are generally accepted beliefs.<sup>9</sup> Premises which he argues for are inferred (either deductively or inductively) from previous premises, which serve as reasons for the truth of the premises that are inferred from them.<sup>10</sup> In such cases, Socrates often gives examples, analogies, or inductive generalizations. These premises give evidence for those premises that are inferred from them, and they tell us where to go to find out whether these premises are indeed true.<sup>11</sup>

According to Santas, that Socrates meets this reasonable expectation is further supported by the fact that he sometimes provides arguments for unsupported premises in elenctic arguments *later* in the dialogue.<sup>12</sup> Moreover, we sometimes find that when Socrates takes a premise for granted in one dialogue, he argues for it in *another* dialogue.<sup>13</sup> For Santas, what this reveals is *Plato’s* decisions about which premises need argument and which premises don’t. Plato’s decisions may be relative to the widely held beliefs of his civilization, among other considerations, and we may question Plato’s choices about which premises he takes for granted, but we would be able to raise questions like these about *any* argument (Santas, “Socratic Method,” pp. 10–11). Nevertheless, it is fairly clear that Plato at least sometimes has Socrates meet the reasonable expectation stated in EA2.

If so, then Santas argues that the problem is still genuine; however, it cannot be applied *a priori* to every elenctic argument where Socrates thinks he establishes that the interlocutor’s initial claim is false. Its application, and thereby scope, would depend on the details of those arguments. More specifically, the problem would apply to those arguments where Socrates takes a premise for granted and it



seems he should not, or where he gives reasons for a premise and those reasons are inadequate.

However, there is at least a way to solve the problem on a case-by-case basis: we can reasonably take Socrates to have given a proof when he uses premises that *as yet* need no justification and argues or gives reasons for premises that need it. If Socrates does these things, and he correctly infers the conclusion from the premises, then we can reasonably think he gives a proof for his conclusion, especially if he later argues for premises he takes for granted in that proof. Given this, it may be that there is no general solution to the problem, but there are at least *piecemeal* solutions, as there are cases where Socrates meets these expectations.

Together, the objections related to each interpretation raise a serious problem for EA and thereby the problem of the elenchos: if EA requires Socrates to give or have particular epistemic reasons for *every* auxiliary premise, then it is open to Kraut's infinite regress objection; if EA requires that Socrates give or have particular epistemic reasons for *some* of the premises, then it is open to Santas' objection that Socrates at least sometimes meets that requirement.

### 3.4 Reply to the Objections

I submit that EA avoids *both* objections and, when properly understood, is justified. To argue this, it is important to briefly revisit the kind of problem that Vlastos had in mind when he raised it. Doing so in light of these two objections will help clarify EA and shed light on a facet of the problem that, to my knowledge, has not been fully appreciated in the scholarship on this problem. I further submit that this facet will provide insight on a new way to adequately solve the problem, which I will sketch in the next and final part of this essay.

To begin with, Vlastos did not intend EA to be interpreted in terms of EA1. Indeed, Vlastos describes the *ad hoc* nature of the agreement to the premises via Socrates' supposed failure to give *particular* reasons for them, and he raises the problem in light of this supposed failure. However, this does not imply that Vlastos expects Socrates to give or have such reasons. Instead, Vlastos interpreted this supposed failure as a symptom of a more general methodological problem that begged for a general solution that explains how the auxiliary premises have a weightier epistemic status than the initial claim.<sup>14</sup>

According to Vlastos, Socrates' elenctic arguments exhibit a *general pattern* whereby Socrates thinks he proves his conclusions from premises for which he secures only the interlocutor's agreement (1994a, pp. 1, 3, 11, 13–14). This gave Vlastos reason to think that this general pattern reflects Socrates' methodological strategy on how he can genuinely refute his interlocutor and argue for his own claims. This also raised the methodological question of how Socrates thinks he and his interlocutor are justified in thinking the auxiliary premises to which they agree are true.<sup>15</sup> This question was posed rather generally to the epistemic rationale behind Socrates' elenctic method; in particular, it is posed to *whatever* auxiliary

premise Socrates might use when practicing his elenctic method, not to particular premises in a particular elenctic argument.

This does not mean that Socrates' reasons in his particular arguments are irrelevant. Indeed, evidence for Socrates' methodological assumptions can be found there, and whatever solution we give to the problem must be tested there. What this does mean, however, is that given the focus of the problem, Socrates' particular arguments are relevant insofar as they shed light on Socrates' epistemic commitments at the methodological level of philosophical inquiry.

Hence, the problem should be understood as a *general* methodological problem that seeks to reveal Socrates' background methodological assumptions regarding the epistemic status of the auxiliary premises, *whatever they may be*. If so, then we should understand the central question of the problem as asking a general methodological question and thereby requiring a general answer. If so, then the problem requires that Socrates must give or have *general* epistemic reasons for the auxiliary premises. We should therefore interpret EA in this way and *not* in the way that is presupposed by EA1, which asks Socrates to give *particular* epistemic reasons for *every* premise he uses. In other words, a more accurate interpretation of EA would be as follows.

EA3 Socrates must give or have *general* epistemic reasons for the truth of the auxiliary premises he uses in his elenctic arguments.

Of course, since these would be general epistemic reasons, they would be applicable to every premise Socrates uses in those cases where he thinks he establishes the falsehood of the interlocutor's initial claim. However, the primary purpose of these reasons would *not* be to secure this or that particular premise; instead, they would provide the general epistemological rationale for how the auxiliary premises have a weightier epistemic status than the interlocutor's initial claim.

Now, one might respond with Santas' solution and argue that to obtain these general assumptions, we need look no further than the fact that Socrates often uses auxiliary premises that *as yet* need no justification and that he often argues or gives reasons for premises that need it. Given this, we can reasonably take Socrates to think that he can give a proof when he meets the more reasonable expectation that he give reasons for *some* the auxiliary premises, namely, those that need such reasons.

This leads me to my next point: Santas is right to point out the argumentative features of Socrates' elenctic practice; however, it seems to me that his piecemeal solution is unsatisfying. The main weakness is that Socrates does *not* emphasize these features in his methodological remarks on how he achieves the aim of establishing truth. In his remarks, he does not say he establishes his conclusions because he uses premises that as yet need no justification and argues for others. To be sure, this is expressed in his elenctic practice, but I think it is important to pay close attention to what Socrates actually says.

In his remarks, Socrates emphasizes the fact that he has secured the interlocutor's *agreement* to the premises and conclusions of his arguments. This is most evident in the *Gorgias*, where the problem of the elenchos is raised most explicitly. There,

Socrates makes several methodological remarks indicating that all he needs is the interlocutor's agreement to accomplish his elenctic aims. In his round with Polus, Socrates distinguishes the oratorical elenchos from his elenctic method: unlike the oratorical elenchos, Socrates' elenctic method achieves the aim of establishing truth, and it does so by securing the interlocutor's sincere agreement with Socrates' position (*Grg.* 471e2–472c5).<sup>16</sup> When Socrates later claims to have refuted Polus, he claims he did so because he just secured Polus' agreement to his conclusions (*Grg.* 475e6–476a2).<sup>17</sup>

When Socrates embarks on his examination of Callicles, he gives a sketch of how they will achieve positive results. In this sketch, Socrates emphasizes that their *mutual agreement* will lay hold of the truth—Socrates does not emphasize any property of the premises he will use, nor does he point to any quality of the arguments he will construct (*Grg.* 486e8–487a1, 487e1–7).<sup>18</sup> This is important because Socrates does *not* emphasize the fact that he uses premises that *as yet* need no justification and that he argues for others.<sup>19</sup> Instead, he emphasizes only the fact that he secures the interlocutor's *agreement* to his conclusions. Since the kind of agreement Socrates seeks is sincere agreement, it therefore seems that, for Socrates, the engine that drives his constructive elenchos is *sincere mutual agreement*.<sup>20</sup>

Given this, Santas' solution is unsatisfying because it does not seem to be Socrates' solution. This is not to say that the features to which Santas points are unimportant; however, it seems they should be understood as secondary to how Socrates understands and explains the constructive efficacy of his elenctic method. Additionally, it seems that Santas was right to point out that Vlastos overstated the claim that Socrates gives “no reasons” for the auxiliary premises.

Vlastos, however, was right to point out that the general rationale underlying Socrates' elenctic method seems to depend on the interlocutor's sincere agreement. He was also right to point out that there is something dreadfully wrong with this methodological dependence. First, even if Socrates were to assume that he establishes the premises on some epistemically adequate basis, he wouldn't know that *the interlocutor* assumes these reasons from the simple fact that he sincerely agrees to the premises, and Socrates rarely checks to ensure that the interlocutor has an epistemically adequate basis for accepting the premises. For all Socrates knows, and as Vlastos points out, the interlocutor can agree to the premises for “epistemically weightless reasons” (1994b, pp. 45–46).

Second, Socrates often dismisses agreement of others as insufficient for thinking a claim is true. As is well known, Socrates argues in the *Crito* (44c7, 48a8) that we should not listen to the opinions of the many. Socrates expresses similar sentiments in the *Laches* (184d5–e7). In the *Gorgias*, Socrates denies that the oratorical elenchos achieves the aim of establishing truth because it uses the opinions of the many to secure the interlocutor's agreement (*Grg.* 471e2–8; 474a2–9). The oratorical elenchos also makes use of false testimony, intimidation, humiliation, and majority vote to secure the interlocutor's agreement, all of which do nothing to secure the truth of that to which they agree (*Grg.* 472a1, 473c1–d5, 473e1–5, 473e7–474a3).

The latter reason sheds light on a facet of the problem that, to my knowledge, has not been fully appreciated in the scholarship: Socrates' methodological dependence

on agreement seems to be *inconsistent* with his own beliefs about the agreement of others. There is something dreadfully wrong with Socrates' methodological dependence on agreement *not* because Socrates fails to give this or that reason for a premise he uses. Instead, it is because he explains the constructive efficacy of his own method in a way that seems inconsistent with his *expressed dismissal* of the agreement of others as deserving any credibility. In light of this, it is entirely legitimate to ask the central question of the problem of the elenchos, for it is raised by the apparent inconsistency between Socrates' methodological dependence on agreement and his own expressed beliefs that deny its reliability.

Therefore, we can ask this more general methodological question because, for Socrates, the constructive efficacy of his elenctic method is grounded on agreement, which he himself seems to reject. Since this inconsistency essentially gives rise to the central question of the problem, and the central question implies that Socrates must give or have general epistemic reasons for the auxiliary premises, then EA is justified when it is understood in this way, EA3. Hence, when properly understood, EA avoids both objections and is justified. Given this, then Vlastos' problem is well-grounded, at least with respect to this assumption understood in this way.

### 3.5 New Strategy for Adequately Solving the Problem

The foregoing, I submit, provides us with some insight on how this problem could be adequately solved. If I am right in what I've been arguing so far, then the problem assumes EA3 in light of two considerations.

- C1 Socrates' stated dismissal of the agreement of others as a reason for thinking a claim is true.
- C2 Socrates' stated methodological dependence on sincere mutual agreement to accomplish the elenctic aim of establishing truth.

Given C1, then it seems Socrates cannot hold C2, for the constructive efficacy of his elenctic method would depend on a feature that he *himself* dismisses. If Socrates thinks agreement of others is inadequate for thinking a claim is true, then it seems he should think the same regarding the agreement between him and the interlocutor—in both cases, it's *mere* agreement. Indeed, this is how Vlastos and other scholars interpreted the kind of agreement that Socrates secures from his interlocutor.

We might say that the difference between the kind of agreement he dismisses and the kind he accepts is that the latter kind is *sincere*; however, it is not clear why Socrates should dismiss the agreement of others as insincere; indeed, it seems that many would have sincerely agreed with Polus that, say, Archelaus is happy (*Grg.* 471a3-d2). It therefore seems that Socrates needs some *other* epistemic reason for thinking that the premises are true; however, since Socrates holds C2, then this strategy for a solution cannot adequately solve the problem.<sup>21</sup>

If so, then a general solution to the problem might lie in solving the apparent inconsistency between C1 and C2—an inconsistency that is crying for a distinction. If so, then we must more closely examine the nature of the kind of agreement Socrates secures in his elenctic practice and distinguish it from the kind of agreement he dismisses as lacking credibility. If we can make this distinction, and then give a plausible explanation for how Socrates can think the mutual agreement between him and the interlocutor lends at least *some* epistemic weight to the auxiliary premises, then this would solve the general methodological problem Vlastos had in mind. If Socrates thought their mutual agreement provided at least some epistemic weight to the auxiliary premises, then this would provide the requisite epistemic distinction Socrates needs to tip the scale in favor of the premises and against the initial claim. This is because Socrates and the interlocutor mutually agree to the auxiliary premises but *disagree* about the initial claim; hence, the auxiliary premises would have justification that the initial claim does not have. Socrates and the interlocutor would therefore have the requisite justification to retain the auxiliary premises and abandon the initial claim.

As I have outlined elsewhere (Santana 2003, pp. 153–197) and will *briefly sketch* here, I think there is evidence to think this strategy will work. First, there is reason to think that we can resolve the inconsistency between C1 and C2. In the cases within the scope of the problem, Socrates secures the interlocutor's agreement to the auxiliary premises under two conditions: (1) *they both must aim to acquire truth and avoid error*; and (2) *they both must say what they really think*.<sup>22</sup> These two conditions ensure that both participants in elenctic inquiry express their considered views on the topic under discussion. They both aim to get at truth and they say what they think is true.

In the kind of agreement Socrates dismisses in C1, these conditions are not established or, at least, it is unclear whether or not they are established. As mentioned above, the agreement of others can be secured by a variety of rhetorical methods and for a variety of aims, none of which may be relevant to the truth of that to which they agree. For this reason, Socrates dismisses this kind of agreement: people say all kinds of things for all kinds of reasons that may or may not be relevant to acquiring truth and avoiding error. Socrates and the interlocutor, however, must agree under the two conditions of elenctic inquiry, which ignores these irrelevant methods and aims. For this reason, we may distinguish the kind of agreement Socrates secures in his elenctic arguments from the *mere* agreement of others.

Second, there is evidence that under these conditions, Socrates thinks the mutual agreement between him and the interlocutor gives them *justification* for the auxiliary premises. To begin with, Socrates' methodological remarks in the *Gorgias* indicate that he thinks their mutual agreement will lay hold of the truth (*Grg.* 486e8–487a1, 487e1–7). After his methodological remarks to Callicles, Socrates explicitly discloses that their mutual agreement gives him *corroboration* or confirmation that what he thinks is true.<sup>23</sup> Given this, it makes sense to think that Socrates' agreement with Callicles gives *Callicles* corroboration that what he thinks is true. Hence, their mutual agreement gives *both* Socrates *and* the interlocutor corroboration that what they think is true (*Grg.* 489a5).<sup>24</sup> If so, then Socrates thinks that agreement

under these conditions secures from the interlocutor something more epistemically weighty than *mere* agreement. He secures *corroborative agreement*, which gives both him and the interlocutor *justification* that the auxiliary premises are true: the interlocutor's sincere agreement gives Socrates justification that his judgments are correct, and Socrates' sincere agreement gives the interlocutor justification that his judgments are correct. By their corroborative agreement, both Socrates and the interlocutor get assurance that their judgments have objectivity.

How? And how is corroborative agreement epistemically weightier than the mere agreement of the many? The fundamental epistemic intuition seems to be that our agreement under the conditions of elenctic inquiry amounts to *something* regarding justification for that upon which we agree; we can't expect to make any philosophical progress at all unless we thought that our agreements under these conditions stood for *something*. This justification is akin to the kind of justification we get when we check our opinions with others.

Suppose you and I see a woman in the distance, and we aim to correctly identify who she is. Sincerely believing that she is Jones, I ask you, "Is that Jones?" If you sincerely think that she is Jones, and you tell me so, then you have given me more reason for thinking that my belief is correct. And the fact that *I* think she is Jones gives *you* more reason to think that *your* belief is correct. Indeed, this kind of justification is weak; and it might be that we're both wrong; nevertheless, our sincere agreement gives us *more* reason to think that our belief is correct than if we were thinking alone. This, I submit, is the kind of justification that Socrates thinks he secures by corroborative agreement.

Suppose, however, that I had reason to doubt that you aimed to identify her correctly, or I had reason to doubt you were being sincere. Even if you agreed that she was Jones, I would be no better off than I would be if I were thinking alone, because I would have no way of telling whether you were sincerely expressing what you thought to be true. This, I submit, is how Socrates generally views the mere agreement of the many: for him, the many agree to lots of things for all kinds of reasons; therefore, it is an open question whether their agreement expresses their sincere beliefs about what they think is true. For this reason, the agreement of the many is dismissible when one merely appeals to their agreement as the basis for thinking a claim is true.<sup>25</sup>

We therefore have reason to think that C1 and C2 are consistent. Socrates dismisses the agreement of others because it is not secured under the conditions of elenctic inquiry or at least it is unclear whether or not such conditions were established. It therefore can be dismissed as epistemically weightless—*mere* agreement. Socrates, however, accepts and even grounds his elenctic method on the agreement between him and the interlocutor because it is secured under the conditions of elenctic inquiry. Such agreement is *not* mere agreement; it is corroborative agreement. Unlike mere agreement, corroborative agreement is epistemically weighty because it provides at least *some* confirmation for what one thinks.

When considered within the context of the problem of the elenchos, corroborative agreement gives both Socrates and the interlocutor the requisite epistemic distinction to solve the problem. Within the context of a single elenctic episode, Socrates

and the interlocutor agree to the auxiliary premises; hence, they have corroboration and thereby justification for them. However, they *disagree* about the initial claim; hence, they *do not* have corroboration for it and thereby *do not* have justification for it. Given this, Socrates and the interlocutor have the requisite justification for retaining the auxiliary premises and abandoning the initial claim, which solves the general methodological problem.

Again, this is just a sketch, and needless to say, this sketch must answer many more questions.<sup>26</sup> However, if it works, then it would offer several advantages, of which I will outline a few. First, it would provide both Socrates and the interlocutor with a *common* reason for retaining the auxiliary premises and abandoning the initial claim. Both Socrates and the interlocutor may have their *own* reasons for thinking that a particular premise is true; however, since those reasons are usually undisclosed it is not clear what they are or whether they are legitimate. Corroborative agreement gives both Socrates and the interlocutor a form of justification that enables them to set these questions aside, for it gives them a common reason that enables them carry out the argument.

Second, it would give both Socrates and the interlocutor a method grounded on an epistemic intuition that hardly needs defense. The epistemic intuition is that agreement by the participants under the conditions of elenctic inquiry offers at least *some* justification for that to which they agree. Indeed, this form of justification would be weak and the conclusions it established would be tenuous. However, this would enable Socrates to avoid defending the epistemological ground for his method, which would enable him to focus on his primary interest in ethics. It would also give Socrates a method grounded on an *accessible* epistemic intuition, which would make his method one that could be used on anyone, understood by anyone, and practiced by anyone.

Third, it would be a solution that is consistent with Socrates' elenctic practice, as indicated by both Kraut and Santas. Socrates can ground his elenctic method on this weak form of justification and nonetheless give compelling arguments by using "eminently reasonable" (Kraut 1983, p. 65) premises or providing arguments for some of the premises, either within that argument or in a later one. So although the epistemic ground for the method is weak, Socrates can nonetheless give arguments of "iron and adamant" (*Grg.* 509a), for the strength of the argument will depend on the strength of the agreements that forge it.

Finally, it would provide a general solution that is consistent with Socrates' own methodological remarks; hence, this solution would be consistent with how Socrates himself understands and explains the constructive efficacy of his elenctic method in the early dialogues—especially the *Gorgias*, where the problem is raised most explicitly.

## Notes

1. Vlastos 1994a, p. 21; my italics. By "Socrates," I refer only to the character in Plato's early dialogues. Unless otherwise stated, all translations are from the Cooper (1997) edition of Plato's complete works.



2. In regards to the former controversy, different solutions have been offered by Vlastos (1994a), Kraut (1983), Bolton (1992), Davidson (1992), Brickhouse and Smith (1994), Irwin (1995), and others. To my knowledge, none of these solutions have gained general acceptance. In regards to the latter controversy, problems have been raised on several grounds. For example, Benson denies the problem because it makes the constructivist assumption that Socrates thinks that a single elenctic episode can *establish* that the interlocutor's initial claim is false. He denies this assumption because it implies that Socrates himself *believes* the premises he uses in his elenctic arguments, and Benson denies this implication (Benson 2000, pp. 54–55, 80). I think that in addition to the epistemic assumption taken up in this paper, the constructivist assumption is also justified (see n. 3). By this, I do *not* claim that Socrates thinks he arrives at constructive conclusions in *every* elenctic argument he gives; rather, I claim that he thinks this in at least *some* arguments.
3. The constructivist assumption (1) implies two others: (2) Socrates thinks his elenctic method *tests* the truth of the initial claim, and (3) Socrates himself *believes* the auxiliary premises that he uses in his elenctic arguments. Each assumption has been hotly debated since Vlastos raised the problem.

Nevertheless, I think there is good evidence to think these assumptions are justified, but since it is beyond the scope of this paper to fully justify these assumptions, I will only point to *Crito* 46b1–48a9 as evidence for my view. There, Socrates addresses Crito's claim that "one must also pay attention to the opinion of the majority" (*Cri.* 44d1). As I read the passage, Crito *explicitly* states quite generally that they should listen to majority opinion. The fact that he is saying this within the context of arguing that it would be right for Socrates to escape prison naturally leads one to suppose that Crito *implicitly* claims that they should listen to majority opinion about just and unjust action. To this, Socrates responds as if he is revisiting a topic that they discussed before and concluded that they should not listen to majority opinion. Socrates gladly revisits the argument but is skeptical that he will change his mind, as Crito thinks he should. Here, Socrates and Crito make contrary claims, and Socrates remarks that they should conduct an examination to decide the issue in a way that will justify *abandoning* the refuted claim and *accepting* the contrary claim. He says, "And I wish to investigate, Crito, in common with you, and see whether our former argument seems different to me under our present conditions, or the same, and whether we shall *give it up or be guided by it* (*easomen chairein ê peisometha auto(i)*)" (*Cri.* 46d7, Fowler transl., *my italics*).

Socrates then proceeds to argue in his usual way, eliciting Crito's agreement to a number of other claims. Socrates then concludes, "We should not then think so much of what the majority will say about us, but what he will say who understands justice and injustice. . . ." (*Cri.* 48a5). Given Crito's explicit and implicit claim, it seems natural that Socrates would argue in this way. Socrates' argument refutes Crito's *explicit* general claim by exposing the falsehood of the *implicit* claim that it entailed.

After his elenctic argument, Socrates clearly thinks he established the falsehood of Crito's claim, for he concludes the argument by saying, "And so you introduced the discussion wrongly [*ouk orthôs*] in the first place, when you began by saying that we ought to consider the opinion of the multitude. . ." (*Cri.* 48a7; Fowler transl.). Here, Socrates clearly expects Crito to give up his initial claim, for he scolds Crito for introducing the idea in the first place. This expectation and scolding makes sense only if Socrates thought that he proved the initial claim false. If Socrates thought that he only revealed a contradiction on Crito's beliefs, then there would as yet be no reason to scold him, for there would be as yet no reason to think that the introduction of majority opinion was a mistake. That Socrates thinks he proves Crito's claim is false is further supported by the fact that Socrates' conclusion provides part of the basis for his decision to stay in prison and be executed. Here, it is quite clear that what Socrates says and does in this passage makes sense only if Socrates thinks his elenchos established that Crito's initial claim is false.

We could therefore think that assumption (2) is justified because Socrates is clearly using his elenctic method to test whether Crito's claim is true or false (*Cri.* 46d7). Assumption (1) is justified because Socrates clearly expects Crito to *give up* his initial claim and accept Socrates' claim that they should not listen to majority opinion (*Cri.* 48a7). Since (1) is



justified, then (3) follows. In addition, I don't see any compelling reason in this passage to deny that Socrates himself *believes* the premises he uses, especially since this argument provides part of the basis for Socrates to deny Crito's plea for Socrates to escape prison.

So this passage alone justifies all three assumptions. This passage, however, does not imply that every elenctic episode is constructive, nor does it imply that Socrates always uses his elenctic method to test for truth, nor does this imply that Socrates always believes the premises he uses.

4. Vlastos' representation of the "standard" elenchos has itself been the subject of controversy. For example, Carpenter and Polansky (2002, pp. 99–100) have raised serious doubts about whether a general account of the elenchos can even be given. Brickhouse and Smith take this a step further by denying that there can be a solution to the problem of the Socratic elenchos on grounds that a definitive account of the elenchos cannot be given in the first place. According to Brickhouse and Smith (2002), "the very idea of 'the Socratic elenchos'—and thus the notion that there is some very special 'problem of the elenchus'—is an artifact of modern scholarship" (Brickhouse and Smith 2002, p. 147).

I think that Carpenter and Polansky (2002, p. 90) provide good evidence that "it is perhaps quite unlikely that there is a single way to conceive each and every elenctic refutation." However, I find Brickhouse and Smith's further suggestion unsatisfying. In my view, neither the problem of the elenchos nor its solution depends on whether a definitive account can be given. There can be a variety of ways in which Socrates makes use of his elenchos, yet there can still be a problem with one of those uses. What the problem requires is there be passages in the early dialogues that correspond to a general pattern in which Socrates thinks he establishes the falsehood of the interlocutor's initial claim via his sincere agreement to the auxiliary premises.

If there are such passages, then the problem and its possible solution arise despite there not being a single way of interpreting every elenctic passage. In my view, there are many such passages, one of which is at *Crito* 46b1–48a9 (see n. 3). In this passage, Crito makes the initial claim that they should listen to majority opinion; Socrates elicits Crito's agreement to several statements that imply the negation of his initial claim; and Socrates expects Crito to give up his initial claim. Whether or not this general pattern is definitive of what we could call *the* Socratic elenchos strikes me as beside the point. Also beside the point is the fact that some elenctic arguments do not follow this general pattern, for such cases would simply be construed as outside the scope of the problem, and any solution we gave would not be applicable to those arguments. Therefore, even if Carpenter and Polansky are correct, there can still be a problem so long as there are elenctic passages that follow the general pattern that give rise to it.

5. Indeed, Vlastos' "standard" elenchos has problems. For example, step (1) isn't wholly true: Socrates sometimes is non-committal about the truth value of the initial claim (*Rep.* I 339b2–6). In any case, we don't see Socrates agree to the interlocutor's initial claim, although he sometimes makes ironic suggestions that what the interlocutor says is true. Kraut (1983, pp. 59–60) raises the following question regarding step (3): Is the elenchos primarily a *deductive* method? I think so, although I found one instance (*Chrm.* 167b1–168e7) where Socrates seems to argue inductively to raise doubts about his interlocutor's initial claim. This does not mean that Socrates uses *only* deductive arguments in his elenctic method. Socrates often argues inductively for auxiliary premises; however, these premises are established in an elenctic argument whose overall structure is deductive. Step (3a) is not an explicit part of Vlastos' "standard" elenchos but it makes explicit what step (3) implies.
6. Benson was right to point this out (2000, p. 34). Benson calls this the "alethic distinction"; however, it doesn't seem to me to be different than the epistemic distinction I am suggesting here. On both conceptions, the point is essentially the same: Socrates must somehow show that the auxiliary premises have a plausibility that the initial claim does not have.
7. For example, in the *Euthyphro*, Socrates relies on Euthyphro's agreement to his assumptions about the Forms and how to construct adequate definitions of them (*Euphr.* 5d1–4, 6d7–e1). In the *Charmides*, he relies on Charmides' agreement that modesty is not a good mate for

- a needy man (*Chrm.* 161a4). In the *Gorgias*, Socrates relies on Polus' agreement that doing injustice is more shameful than suffering it (*Grg.* 474c6–8).
8. Santas, "Socratic Method" (unpublished draft cited with the author's permission), pp. 4, 5, 10, 17.
  9. *Ibid.*, p. 13. Santas notes that in *Republic* I, Socrates assumes (and the interlocutor agrees) that when horses are harmed they are made worse horses, and worse with respect to the excellence of horses (*Rep.* I 335b5). He also assumes that it is not the function of heat to cool, nor is it the function of dryness to moisten (*Rep.* I 335d2–4). Later, Socrates assumes that rulers are fallible (*Rep.* I 339c2).
  10. Santas, "Socratic Method," p. 6. In *Republic* I, Santas closely examines Socrates' refutation of Polemarchus' definition that justice is helping friends who are good and harming enemies who are bad. At the end of this argument, Socrates claims to have shown it false. Within the argument, however, he gives at least three inductive analogies to establish three different premises, one of which is that harming human beings makes them worse with respect to human virtue (*Rep.* I 335b5-d9). According to Santas, this argument shows quite clearly that it is simply not true that Socrates does not argue for the premises. Rather, he takes some premises for granted and argues for others (Santas, "Socratic Method," pp. 8–11).
  11. *Ibid.*, p. 11. For example, Santas notes, we could check with experts on horsemanship to determine whether harming horses makes them worse with respect to the excellence of horses, or we could check with physicists to determine whether heat can cool.
  12. For example, in his refutation of the definition that justice is helping friends who are good and harming enemies who are bad, Socrates takes for granted that justice is human virtue (*Rep.* I 335c3). Later in the dialogue, Thrasymachus denies this claim, and Socrates later argues for it (*Rep.* I 348c7, 351c6–352d3).
  13. For example, in the *Laches*, Socrates secures the interlocutor's agreement that courage is *kalon*; in the *Charmides*, he secures the interlocutor's agreement that temperance is *kalon*. In both instances, he doesn't offer any reasons for these assumptions. In the *Gorgias*, however, Callicles challenges these assumptions by denying that justice is *kalon*, and Socrates gives reasons for this assumption. As Irwin points out (1995, pp. 48–51, 124–5), and Santas agrees, we can see this as an instance where Plato offers reasons for assumptions he makes in earlier dialogues.
  14. There are several reasons to think this is the case. First, Vlastos makes this point in the opening lines of his essay (1994a, p. 1; *my italics*).

In Plato's earlier dialogues. . . Socrates' inquiries display a *pattern of investigation whose rationale he does not investigate*. They are constrained by rules he does not undertake to justify. . . . He never troubles to say why this *way of searching* is the way to discover truth or even what this way of searching is.

Here, it is clear that Vlastos focus is in Socrates' general rationale for how his elenctic method discovers truth. Second, when Vlastos searches for possible epistemic reasons to ground Socrates' elenctic method, he points out that Socrates does not rely on two kinds of reasons that Aristotle used to ground his methods of demonstration and dialectic: self-certifying truths and *ta endoxa*, respectively (*Top* I 100a27-b23). These are general epistemic reasons that primarily serve to ground their respective methods, but could also be used to ground particular premises. The fact that Vlastos denies these as possible candidates is evidence that he was looking for reasons with this level of generality. Third, Vlastos solves the problem by attributing to Socrates several general assumptions (1994a, pp. 25–28). Fourth, Vlastos highlights the generality of the problem when he explains how his solution relates to Plato. Vlastos states (1994a, p. 29):

The question, "How could this be true?", which never disturbed Socrates, could hardly help disturbing Plato when he writes the *Gorgias*. For here he puts into Socrates' mouth that flock of *obiter dicta* which reveal the assumptions on which he predicates his confidence that the elenctic method establishes truth and falsehood.

Here, Vlastos places the question of how a premise could be true within the context of the general methodological problem. Hence, for Vlastos, the problem of the Socratic elenchos is a general one that requires a general solution.

15. This methodological point has been generally recognized in the scholarship on this problem (Kraut 1983, p. 61; Polansky 1985, p. 253; Benson 2000, pp. 34–5; Brickhouse and Smith 1994, pp. 10–11; Genzler 1994, pp. 268–9); however, what has not been made explicit is what this implies about how EA should be interpreted and what implications it has for an adequate solution to the problem.
16. S: My wonderful man, you're trying to refute me in oratorical style, the way people in law courts do when they think they're refuting some claim. There, too, one side thinks it's refuting the other when it produces many reputable witnesses on behalf of the arguments it presents, while the person who asserts the opposite produces only one witness, or none at all. This "refutation" is worthless, as far as truth is concerned, for it might happen sometimes that an individual is brought down by the false testimony of many reputable people. Now too, nearly every Athenian and alien will take your side on the things you're saying. ... Nevertheless, though I'm only one person, I don't agree with you. You don't compel me; instead you produce many false witnesses against me and try to banish me from my property, the truth. *For my part, if I don't produce you as a single witness to agree with what I'm saying, then I suppose I've achieved nothing worth mentioning concerning the things we've been discussing.* And I suppose you haven't either, if I don't testify on your side, though I'm just one person, and you disregard all these other people. (My italics.)
17. S: So you see, Polus, that when one refutation is compared with the other, there is no resemblance at all. Whereas everyone but me agrees with you, you are all I need, although you're just a party of one, for your agreement and testimony. *It's you alone whom I call on for a vote; the others I disregard.* (My italics.)
18. S: I know well that if you concur with what my soul believes, then that is the very truth. . . It's clear, then, that this is how these matters stand at the moment. If there's any point in our discussions on which you agree with me, then that point will have been adequately put to the test by you and me, and it will not be necessary to put it to any further test, for you'd never have conceded the point through lack of wisdom or excess of shame, and you wouldn't do so by lying to me, either. You are my friend, as you yourself say, too. *So our mutual agreement will really lay hold of truth in the end [hê emê kai hê sê homologia telos edê hexei tês alêtheias].* (My italics)
19. Indeed, Socrates claims to have given arguments of "iron and adamant" (Grg. 509a) but he doesn't point to this as a reason for how he accomplishes his elenctic aim of establishing truth.
20. Socrates also suggests this in *Republic* I (348a1-b7):

S: If we oppose him with a parallel speech about the blessings of the just life, and then he replies, and then we do, we'd have to count and measure the good things mentioned on each side, and we'd need a jury to decide the case. But if, on the other hand, we investigate the question, as we've been doing, by seeking agreement with each other [*arti an homologoumenoi pros allêlous skopômen*], we ourselves can be both jury and advocates at once.

21. This route has been taken by several scholars. For example, some have argued along the lines delineated by Kraut and Santas. Others have argued that the premises he uses are generally accepted beliefs (Polansky 1985, pp. 249–53). Still others have argued that Socrates gains inductive evidence for the truth of his conclusions by the fact that he repeatedly secures the interlocutor's agreement to them (Brickhouse and Smith 1994, pp. 18–20).

The main weakness of the first two kinds of solutions is that Socrates does not offer such general epistemological reasons to ground the rationale behind his elenctic method; instead, he seems to focus on C2. Indeed, Socrates can and does use generally accepted beliefs as

premises in his elenctic arguments; however, as Vlastos notes (1994a, pp. 15–6), the mere fact that a premise is generally accepted cannot by itself be a reason for thinking it is true.

The main weakness of the third kind of solution is that if we hold Socrates to C1, then it seems that the interlocutor's agreement to his conclusions should not by itself provide *any* inductive evidence for thinking that they are true. If Socrates dismisses the agreement of others as lacking credibility, then it is unlikely that he should think that *repeated* agreement should be any better.

22. Evidence for the first condition can be found at *Grg.* 457c5–458b5. There, Socrates and Gorgias establish this primary aim early in the dialogue and it is from within the context of this aim that the rest of the dialogue plays out.
23. S: Now, isn't it a rule of the many that it's just to have an equal share and that doing what's unjust is more shameful than suffering it, as you yourself were saying just now? Is this so or not? Be careful that you in your turn don't get caught being ashamed now. Do the many observe or do they not observe the rule that it's just to have an equal and not a greater share, and that doing what's unjust is more shameful than suffering it? Don't grudge me your answer to this, Callicles, so that if you *agree* with me I may have my *confirmation* from you [*ean moi homologêsê(i)s bebaiôsômai êdê para sou*], seeing that it's the agreement of a man competent to pass judgment. (*Grg.* 488e8–489a7; my italics.)
24. Socrates makes similar remarks at *Protagoras* 348c6–e2:

S: "Protagoras," I said, "I don't want you to think that my motive for talking with you is anything else than to take a good hard look at things that continually perplex me. I think that Homer said it all in the line,

*Going in tandem, one perceives before the other.*

Human beings are simply more resourceful this way in action, speech, and thought. If someone has a private perception, he immediately starts going around and looking until he finds somebody he can show it to and have it corroborated [*bebaiôsêtai*]. And there is a particular reason why I would rather talk with you than anyone else: I think you are the best qualified to investigate the sort of things that decent and respectable individuals ought to examine, and virtue especially."

25. Incidentally, Socrates does not always dismiss the views of the many; indeed, he sometimes takes their views quite seriously. However, it is important note that he does so when one could reasonably think that the many would express these beliefs under the two conditions of elenctic inquiry. For this reason, among others, we see Socrates take up the view of the many on *akrasia* in the *Protagoras*.
26. For example, one could concede that corroborative agreement may render justification for that to which Socrates and the interlocutor agree, but one might nonetheless object that it does not render truth. Two people could have justification for believing something and still be wrong. Socrates, however, aimed also to acquire truth and avoid error. If so, then it would seem that corroborative agreement gives us, at best, only half the story on how the Socrates' elenctic method establishes truth. This important point has been raised to me by Rod Jenks, and I agree that a fully adequate solution must also inquire into Socrates' methodological presuppositions about truth.

## Bibliography

- Benson, H. 2000. *Socratic Wisdom: The Model of Knowledge in Plato's Early Dialogues*. New York: Oxford University Press.

- Benson, H. 2002. "Problems with Socratic Method." In *Does Socrates Have a Method? Rethinking the Elenchus in Plato's Dialogues and Beyond*, edited by G.A. Scott. University Park: Pennsylvania State University Press.
- Bolton, R. 1992. "Aristotle's Solution to the Problem of the Socratic Elenchus." In *The Philosophy of Socrates*, edited by K.J. Boudouris. Vol. 2. Athens: Kardamitsa Publishing.
- Brickhouse, T., and N.D. Smith. 1994. *Plato's Socrates*. New York: Oxford University Press.
- Brickhouse, T., and N.D. Smith. 2002. "The Socratic Elenchos?" In edited by G.A. Scott.
- Burnyeat, M., ed. 1994. *Socratic Studies*. New York: Cambridge University Press.
- Carpenter, M., and Polansky, R. 2002. "Variety of Socratic Elenchi." In edited by G.A. Scott.
- Cooper, J., ed. 1997. *Plato: Complete Works*. Indianapolis: Hackett Publishing.
- Davidson, D. 1992. "The Socratic Conception of Truth." In *The Philosophy of Socrates*, edited by K.J. Boudouris. Vol. 2. Athens: Kardamitsa Publishing.
- Genzler, J. 1994. "Recollection and the Problem of the Socratic Elenchus." *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, 257–95.
- Irwin, T. 1995. *Plato's Ethics*. Oxford: Oxford University Press.
- Kraut, R. 1983. "Comments on Gregory Vlastos, 'The Socratic Elenchus'." *Oxford Studies in Ancient Philosophy* 1:59–70.
- Polansky, R. 1985. "Professor Vlastos' Analysis of the Socratic Elenchus." *Oxford Studies in Ancient Philosophy* 3:247–60.
- Santana, A. 2003. *The Problem of the Socratic Elenchus*. Ann Arbor: UMI.
- Santas, G. unpublished. "Socratic Method in Plato's *Republic* I." Unpublished draft cited with the author's permission, 1–18.
- Scott, G.A., ed. 2002. *Does Socrates Have a Method? Rethinking the Elenchus in Plato's Dialogues and Beyond*. University Park: Pennsylvania State University Press.
- Vlastos, G. 1994a. "The Socratic Elenchus: Method is All." In edited by M. Burnyeat.
- Vlastos, G. 1994b. "Socrates' Disavowal of Knowledge." In edited by M. Burnyeat.

## Chapter 4

# Santas, Socrates, and Induction

Mark L. McPherran

The publication of Gerasimos Santas' book *Socrates* in 1979 helped to initiate the resurgence of the study of Socrates that continues from that time through to the present day.<sup>1</sup> Prior to its appearance, for example, there existed only quite general studies of Socratic thought, ranging from A. E. Taylor's venerable 1932 *Socrates* to Norman Gulley's 1968 *The Philosophy of Socrates*. Even subsequent to the publication of Santas' book, the next major work on Socrates to appear—Gregory Vlastos' *Socrates: Ironist and Moral Philosopher* (1990)—was still a decade away.

What impressed so many of Santas' initial readers was his book's insistence that we take the Socrates of *Crito* 46b seriously in all our interpretative approaches to the Socratic dialogues. In that passage, we recall, Socrates famously asserts that

Not now for the first time, but always, I am the sort of man who is persuaded by nothing except the argument (*to(i) logô(i)*) that seems best to me when I reason (*logizomenô(i)*) about the matter. (*Cri.* 46b3–6)

This Socrates is a rational revolutionary who demands that we not be persuaded by *his* arguments and the results of *his* interrogations of others unless we too have worked through the claims at issue in a rigorous and painstaking fashion. By trying to make sense of Socrates in his own terms—as a man devoted to argument—Santas was able to offer a comprehensive and detailed presentation of Socrates' political philosophy (Part One of his book) and moral theory (Part Three). The result was a text that made clear as never before the richness, complexity, and—above all—thoroughgoing *rationality* of Socratic philosophy.

All that would be accomplishment enough, but in perhaps the most innovative and seminal section of the book—Part Two, on Socratic Method—Santas offers a comprehensive account of *how* Socrates conducted his argumentative examinations of others and *how* he arrived at his own distinctive views. In order to adumbrate this account of Socrates' philosophical method, Santas recalls Aristotle's summation of

---

M.L. McPherran (✉)  
Simon Fraser University, Burnaby, BC, Canada  
e-mail: mark\_mcpherran@sfu.ca

Socrates' distinctive contributions to philosophical method (Santas 1979, p. 97):

Socrates occupied himself with the excellences of character, and in connection with them became the first to raise the problem of universal definitions . . . . But it was natural that Socrates should seek the essence. For he was seeking to deduce, and the essence is the starting-point of deductions . . . . For two things may be justly ascribed to Socrates—inductive/epagógic arguments (*epaktikoi logoi*) and universal definition, both of which are concerned with the starting-point of science. But Socrates did not make the universals or the definitions exist apart; his successors, however, gave them separate existence, and this was the kind of thing they called Ideas. (*Met* M.4 1078b7–32; transl. after Jonathan Barnes.)

I think Aristotle is best understood to be claiming here not that Socrates was the first to engage in or to conceive of inductive/epagógic argumentation (see below for the distinction between inductive and epagógic argument) and the pursuit of definitions in a philosophical context, but that it was Socrates who first made the quest for definitions central to philosophical inquiry and, in addition, made frequent, systematic use of *epagógê* in his elenctic investigation of such definitions.<sup>2</sup> Plato and Xenophon also appear to target *epagógê* as an innovative, distinguishing mark of Socratic methodology when they have Socrates' interlocutors complain that Socrates prattles on far too much about "his favorite topic" (*Memorabilia* 1.2.37)—blacksmiths, cobblers, cooks, physicians, and other such tiresome, menial craftspeople—in order to generate and test general principles concerning the alleged craft of virtue (*Smp.* 221d–222a; *Grg.* 490c–491c; Xenophon, *Memorabilia* 1.2.33–7). It is remarkable, then, that even today a search of the online *Philosopher's Index* reveals not a single article addressing the subject of Socratic induction or *epagógê*. One might begin to worry that this is because there is little to be said usefully on the topic, were it not for the chapter on *epagógê* found in Richard Robinson's 1953 book, *Plato's Earlier Dialectic*, followed by Santas' own invigorating study.<sup>3</sup> However, Robinson and Santas, and others such as Gregory Vlastos, offer us conflicting accounts of epagógic argumentation. For example, Santas, unlike Robinson and Vlastos, finds many legitimate instances of probabilistic inductive *epagógê* in the early dialogues (Santas 1979, pp. 136–55).

Here I want to argue that Santas was correct in his account, and correct in a way that ought to be better appreciated than it is. But, additionally, his insights can benefit from a few amendments. The point of offering both an appreciation and an emendation is not an idle one, for it seems to me that the secondary literature continues to be littered with cases where the interpretation of a particular Socratic *elenchos* is left problematic by its failure to appreciate the role *epagógê* plays in securing the interlocutor's assent to a crucial premise.<sup>4</sup> Hence, there is clearly a need for a fresh, critical account of Socratic *epagógê*—one tied to its occurrences in several key Socratic *elenchoi*. This paper is an introductory sketch of just such an account, taking the work of Santas as its springboard. Among other things, I will show that not only does Socrates sometimes use a non-inferential epagógic form to *communicate* the meaning of some general claim in order to obtain or cement an interlocutors' consent to that claim ("explicative *epagógê*"—what Gregory Vlastos called "intuitive induction," see below), but that *pace* some commentators, and in validation



of Santas, Socrates also uses another, legitimately inductive form of *epagôgê* to persuade interlocutors to adopt various beliefs. I will discuss the implications this has for the current controversy as to whether Socrates uses the *elenchos* to somehow warrant truths (constructivism) besides using it to test his interlocutors for the consistency of their beliefs and the legitimacy of their related knowledge-claims (see below).

#### 4.1 *Epagôgê* and Induction

Richard Robinson found three types of *epagôgê* in the Platonic dialogues: (1) arguments from a single proposition or set of coordinate propositions that serve as premises to another proposition superordinate to the premise set—for example, “the opposite of beauty is ugliness and only ugliness, the opposite of goodness is badness and only badness; therefore, every opposite has an opposite, and only one” (cf., e.g., *Prt.* 332c–d); (2) arguments from a single proposition or set of coordinate propositions that serve as premises to another coordinate proposition—for example, “expert pilots have the best success at sea-faring; therefore, expert builders have the best success at building” (cf., e.g., *Ap.* 27b; *H. Ma.* 284a–b); and (3) arguments from a single proposition or set of coordinate propositions that serve as premises to another proposition superordinate to the premise set followed by an inference back to a subordinate proposition—for example, “men are mortal, thus human beings are mortal, and thus women are mortal” (cf., e.g., *Euphr.* 10a–d; Robinson 1953, pp. 35–8.). For type 2 arguments, and for type 3 arguments in those cases where the superordinate claim goes unmentioned, the mode of reasoning should be identified as that of inductive analogy.<sup>5</sup> In his own analysis, as we will see, Santas treats Socrates’ inductive analogies in a separate section of his book (ch. V, sec. 3). Next, Robinson posited three conceptions of epagogic movement to a universal that types 1 and 3 (and 2 implicitly) might employ (none of which are explicitly distinguished by Socrates or Plato): (A) Intuition of the Universal through a sample of coordinate cases, “thus obtaining certainty” (e.g., *Euphr.* 10a–d; *Prt.* 332c), (B) Complete Enumeration of a set of coordinate cases followed by trivial deductive inference to a universal (e.g., *Grg.* 474d–475b; *Meno* 87e–89a [cf. *An. Pr* II.23 68b28–9: “For *epagôgê* proceeds through an enumeration of all the cases”]), and (C) Probable Inductive Generalization Employing a Survey of Coordinate Cases that lead to a probable generalization that may be overturned by the discovery of a disconfirming instance (Robinson 1953, p. 35; cf. Guthrie 1971a, pp. 107–8). In general, and although the relation of epagogic arguments to elenctic ones varies, it is common for a Socratic *elenchos* to employ at least one premise that is secured epagogically.<sup>6</sup> I shall now assess this triad in reverse order, arguing that conception (C) (Probable Inductive Generalization Employing a Survey of Coordinate Cases), modified somewhat by (A) (Intuition of the Universal) gives us the correct account of one sort Socratic *epagôgê*—one that resembles what we would call inductive generalization.



(C) *Probable Inductive Generalization Employing a Survey of Coordinate Cases*

Robinson states that “there seems to be no clear case of the conception of epagoge as merely probable in the dialogues,” and so eliminates inductive generalization from Socrates’ methodological toolbox.<sup>7</sup> This is a rather startling conclusion, since one would naturally think that Socrates would recognize and employ the everyday notion of probability with which he is familiar in many of his arguments when moving from a review of cases to a generalization based on them.<sup>8</sup> Nevertheless, Robinson’s claim may seem true at first glance, given that Socrates never offers any classic inductive generalizations composed of a large and/or stratified number of samples. In fact, he rarely cites more than five coordinate cases in composing an *epagôgê* (Robinson 1953, pp. 33–4), and his use of modal qualifiers signaling probability within them is indeed quite rare (e.g., *Charmides* 159a–160e, on which see below). However, I think that we can explain Socrates’ avoidance of both large-sample inductive generalization and inductive modal qualifiers. First, we may hypothesize that Socrates has confidence in some modified form of inductive generalization that he takes to establish conclusions that are highly probable, indeed, conclusions so probable that in some cases they border on practical certainties. Hence, from Socrates’ point of view, he would have no need for large-sample, inductive generalizations. We will consider examples of this sort below. In addition, the construction and elaboration of lengthy or complex inductions would needlessly delay his central therapeutic concern to test for and uncover his interlocutors’ lack of knowledge using the *elenchos*. What this task requires is that Socrates secure his interlocutors’ assent to the various propositions to which they *already* subscribe in order to reveal the inconsistency of a part of their belief-set. Hence, the issue of whether or not each act of assent is itself adequately justified is entirely secondary—if not entirely irrelevant—depending on the dialogical context.<sup>9</sup> In short, since Socrates appears most interested in *uncovering* an interlocutor’s already existing beliefs rather than providing him or her with new and/or better ones, his elenctic mission does not require extensive inductive argumentation for the truth of any of the premises of an *elenchos*. Thus, the fact that Socrates *does* nevertheless seem to engage in what look to be inductive arguments for the truth of some of the propositions of his *elenchoi* requires explanation—something I will provide below. In any event, since Socrates’ use of the *elenchos* is predicated on the assumption that its conclusions are always open to reexamination (see, e.g., *Euphr.* 15c–d), we can suppose that he regarded many or most of the *epagôgê* that contribute to any *elenchos* as establishing high probabilities/practical certainties, but not actual certainties.<sup>10</sup>

(B) *Complete Enumeration of Instances*

Robinson calls this “the most visible” of the three conceptions of Socratic *epagôgê*, and also “the most useless and impractical” (Robinson 1953, p. 36). This latter assessment seems fair, given the problems involved in making such complete reviews in the context of constructing *elenchoi* in a dialogical arena. But since from our perspective—and Socrates’ as well, we would hope—complete enumeration is

tantamount to deduction rather than argumentation of an inductive sort, it seems odd to hold that such a mode of inference is epagogic rather than deductive—this is especially true in view of Aristotle’s explicit dichotomy between the two kinds of inference found in *Posterior Analytics* I.1 and *Topics* I.12. After all, if one has examined every single physician in Athens and has found each one of them to be snub-nosed and male, then there is little need to invoke the characteristic epagogic formula *kata ton auton logon* (e.g., *Grg.* 460b4)—“by parity of reasoning”—in order to conclude that all Athenian physicians are snub-nosed and male. The larger problem here, however, is that Robinson provides no convincing evidence that Socrates *ever*—let alone frequently—conducted his *elenchoi* on the basis of what he understood to be complete enumerations *as opposed to* Intuitions of the Universal (type A). For example, Robinson cites the text of *Meno* 87e–88d as providing an example of arriving at the conclusion that “nothing is really useful unless it is guided by wisdom” on the basis of complete enumeration, since it begins with the line, “Let us inquire what sort of things benefit us, *taking them up one by one*” (87e5–6; my italics). However, the text following that line—“. . . health we say, and strength, and beauty, and also wealth. We say these things, *and others of the same kind*, benefit us . . .” (87e6–7; my italics)—indicates that a genuinely *complete* enumeration is not being conducted. Again, Robinson may be right that at *Gorgias* 474d–475b Socrates talks *as though* he is conducting a complete enumeration of all cases of *kalon* so as to establish that the *kalon* is always the useful or the pleasant (Robinson 1953, p. 36), but that talk is nevertheless compatible with Socrates thinking that his enumeration is or is possibly incomplete. So it seems to me that we should remove both classic, large-sample inductive generalization and complete enumeration as established ways of interpreting Socratic *epagôgê*. Hence, Intuition of the Universal (A) remains the most promising account, although I shall argue that this interpretation also requires modification (one that is a compromise between [A] Intuition of the Universal and [C] Probable Inductive Generalization Employing a Survey of Coordinate Cases).<sup>11</sup>

(A) *Intuition of the Universal (yielding certainty)*

Robinson has surprisingly little to say in regards to what appears, even on his own reading, to be the most promising account of Socratic *epagôgê*. Gregory Vlastos, on the other hand, endorses the enumeration of cases under this description as constituting the most “distinctive” kind of Socratic *epagôgê*, and so proceeds to offer an account of Intuition of the Universal that explains why it could be thought to yield non-probabilistic certainties. According to Vlastos (Vlastos 1991, pp. 267–8), we find a good example of Intuition-of-the-Universal-type-*epagôgê* at *Ion* 540b–d, which—bypassing the rhetorical features of its dialogical presentation—he represents as follows:

- (P1) The pilot is the one who knows best what should be said to the crew of a storm-tossed ship.
- (P2) The doctor is the one who knows best what should be said to the sick.

- (P3) The cowherd is the one who knows best what should be done to calm down angry cattle.
- (P4) The expert in wool is the one who knows best what should be said to women working wool.
- (P5) The military expert is the one who knows best what the general should say to the troops.
- (C1) If C is a craft then its master is the one who knows best matters falling within its subject matter (So, if C is not—say—the rhapsode’s craft, then it is not the rhapsode, but the master of C, who knows best the matters which fall in the domain of C).

The structure of this particular *epagôgê* resembles Probable Inductive Generalization Employing a Survey of Coordinate Cases (C) because C1 is more general than any of the statements that lead up to it, but it is not truly inductive because the argument does not hold out the possibility that C1 might be falsified by experience. For if, say, we found some putative non-doctor *x* to know medicine as well as or better than some actual doctor *y*, we would simply claim that *x* is—after all—really a doctor—thus saving C1 from the putative counterexample. Accordingly, Vlastos concludes that with this textual instance of what he alleges to be a typical case of Socratic *epagôgê*

...we are not leaving logically open the possibility that there might be some craft  $C_a$  such that the master of some other craft  $C_b$  or a layman who is master of neither might have knowledge of matters falling in the domain of craft  $C_a$  which is superior to that of the master of craft  $C_a$ . Here the truth of the conclusion *is built into the meaning of its critical term* “master of a craft”: anyone who claims to be a master of a given craft but does not possess relevant knowledge superior to that of a master of some other craft or of no craft at all would be *ipso dicto* disqualified as a fake. . . .in Socrates’ epagogic arguments there is “reference to some instances [of a general statement] which *exhibit the meaning of the statement* by exemplifying it, rather than *prove* it. . . .” (Vlastos 1991, p. 268, Vlastos’ italics; cf. Guthrie 1971a, p. 107)

On this account of Socratic methodology, Socrates provides no real empirical support for the theses he puts to his interlocutors (*pace*, e.g., Graham 1990, and Kraut 1983, p. 60). Instead, he uses generalizing *epagôgê* to *communicate* the meaning of some general claim in order to obtain his interlocutors’ consent to that claim (or consent to what Socrates counts as another particular instance of it). The method is only what is sometimes termed “intuitive induction,” which does not actually designate a form of inference.<sup>12</sup> Induction, then, plays virtually no role in Socratic *epagôgê*, or, thus, Socratic *elenchoi*, according to Vlastos.<sup>13</sup> This result might, in turn, be taken to indicate that we should therefore think of Socrates as always or almost always obtaining an interlocutor’s assent to the premises of an *elenchos* not by persuasion but by mere explanation—a form of explanation that helps the interlocutor to recall and state his or her actual beliefs, beliefs held prior to his or her encounter with Socrates.<sup>14</sup>

This raises the issue of whether Socrates *ever* tries to persuade an interlocutor of the truth of a premise epagogically. Let me address the issue by assuming a

commonsense distinction between explanation and persuasion, and belief and non-belief.<sup>15</sup> This move yields a four-part matrix against which to measure any example of Socratic interrogation: (a) Socrates tries to persuade interlocutor L of proposition *p* where L does not believe *p*; (b) Socrates tries to persuade L of *p* where L already believes *p*; (c) Socrates tries to explain *p* to L, where L does not (yet) believe *p*; and (d) Socrates tries to explain *p* to L, where L already (in some sense) believes *p*. Now, it seems clear to me that Vlastos is in essence contending that (1) there are few or no epagogic cases of (a)—Socrates rarely or never tries to persuade some L of proposition *p* epagogically where L does not already believe *p*. An anti-constructivist in respect of the *elenchos* might in turn embrace this view, thinking that (2) if there were in fact any such cases of persuasion, rather than explanation, they would weigh heavily against anti-constructivism (defined as the view that Socrates *only* employs the *elenchos* to test *already existing beliefs* in order to test moral knowledge claims).

But we need not embrace this second claim (2). For it seems as though we can modify anti-constructivism without destroying it by adding to this position the claim that Socrates sometimes reveals belief-inconsistency in interlocutors only *after* he convinces them to adopt a belief or two that they had not previously held (beliefs inconsistent with other, possibly less-well-grounded beliefs they already profess to hold). Naturally, this view demands an account of what might at first glance seem to be a rather brutal form of education—but I would suggest that anti-constructivists have sufficient resources in order to show that this is only an apparent snag. In brief, they can maintain that when Socrates encounters an interlocutor professing expert moral knowledge, and has or begins to have suspicions that he or she does not possess such knowledge, then in certain cases an effective way to test that claim and possibly bring the interlocutor's ignorance to light would be to try to convince him or her of some belief that Socrates suspects will reveal the inconsistency of the interlocutor's belief set. Naturally, we would expect Socrates to argue epagogically for those beliefs that he finds true, and so he cannot be accused of misleading interlocutors with this procedure. Moreover, Socrates may assume that whatever he convinces an interlocutor of epagogically would be a belief that any genuine moral expert would accept, and his or her resistance to such acceptance may in turn help Socrates diagnose any interlocutor's particular psychic malady.<sup>16</sup> Finally, Socrates' interrogation of some interlocutors may also be driven by more general moral concerns. In the case of Euthyphro, for example, we have reason to believe that Socrates desires to test Euthyphro's knowledge-claims concerning piety in order to persuade Euthyphro to abandon what appears to Socrates to be a morally dangerous lawsuit against his father (see McPherran 2003, pp. 9–10; 2001; 2002). If so, then we may expect Socrates to attempt to persuade an interlocutor of those propositions that will prevent him or her from pursuing morally harmful actions (see, e.g., *Cri.* 46b–49e; *Memorabilia* 1.4.1, 4.3.1–18).

Be that as it may, what about Vlastos' claim (1) that Socrates rarely or never tries to persuade some interlocutor L of proposition *p* epagogically? Does Socrates instead—on many occasions—try to convince interlocutors epagogically of some proposition *p per* (a)? Well, Socrates does *say* that he tries to persuade people of various things at *Apology* 30b, 36b, and *Crito* 50e–51c—and then at *Gorgias*

492e–494a we find him telling Callicles that he wants to *persuade* him to change his mind (and Xenophon provides several cases where cross-examination gives way to persuasion [e.g., *Memorabilia* 4.2.8–40], sometimes even of a bold, theological nature [e.g., *Memorabilia*. 1.4.1–19; 4.3.1–18]). *Euthyphro* 7c–d also seems to provide a clear case of epagogic persuasion: here Socrates is portrayed as trying to bring Euthyphro through a consideration of cases to conclude that the gods differ about the same sort of thing that we humans differ about (namely, particular moral assessments). Again, at *Republic* 353b–d, it would be implausible to suppose that Socrates was not trying to persuade, by means of an *epagôgê*, a reluctant Thrasymachus that “anything that has a function performs it well by means of its own peculiar virtue.” Finally, it seems to me that anyone attempting to maintain that there are few or no epagogic cases of type (a)—that Socrates rarely or never attempts to persuade his interlocutors to accept beliefs that they did not harbor prior to their encounter with him (hence, that virtually all cases of *epagôgê* are cases of explicative *epagôgê*: of Socrates leading interlocutors *via* explanation to state what they already believe)—has to address what we may term the “Number of Instances Problem.” That is, if Socrates always or generally uses *epagôgê* to explain but not persuade, why does he frequently use more cases—more than just one or two sample cases—than are needed for mere explanation? The obvious solution to this problem, of course, is to dissolve it by maintaining that the number of examples needed to explain adequately an abstract claim to a particularly dense interlocutor can be quite high. But I find this an unconvincing way of side-stepping the issue in view of a reasonable interpretive assessment of a number of such purported examples in our texts (e.g., *Chrm.* 159b–160d, below).

In any event, it does seem that Vlastos is correct to take *Ion* 540b–d as providing an example of *one sort* of *epagôgê*—explicative *epagôgê*—we can find in Plato and Xenophon (see again, e.g., *Euthyphro* 10a–d). For instance, as Vlastos elsewhere notes, at *Protagoras* 332a–e Socrates uses the three cases of beauty/ugliness, goodness/badness, and high pitch/low pitch to show that “for everything that admits of an opposite there is one opposite and no more” (Vlastos 1956, p. xxix, and n. 18). Here, it seems clear, there is in fact no logical opening for the presentation of a counterexample, for given the very concept of what it is to be an opposite as Socrates would define “opposite,” it is inconceivable that some property *x* might have *two* opposites *y* and *z*. A similar case is provided by the *elenchos* of *Gorgias* 460a–e, which—now following Santas—can be formalized as follows:

- (P1) Someone who has learnt the things pertaining to the craft of carpentry is a carpenter.
- (P2) Someone who has learnt the things pertaining to the craft of music is a musician.
- (P3) Someone who has learnt the things pertaining to the craft of medicine is a doctor.
- (C1) Other cases of learning a craft follow the same principle (*logos*): a person who learns the things pertaining to a craft is such as that craft makes him.
- (P4) Justice is a craft (implicit).

- (C2) Someone who has learnt the things pertaining to the craft of justice is a just person.  
 (P4) A just person does what is just.  
 (P5) A just person of necessity wants to do just things.  
 (C3) A just person will never do injustice.

Commenting on what he terms “the inductive generalization” P1 to C1, Santas says

Despite the fact that only three instances [P1–P3] are considered, and despite the fact that all the instances are from the practical science crafts, the argument is very strong. For the connection between something being a craft-science and the fact that someone who learns its subject matter becomes a successful practitioner of it is not simply conjunction; it is an evidential connection. . . . The argument to C1, then, is strong and sound. (Santas 1979, p. 151)

Here, I think, Vlastos would be right to object that the “argument” for C1 is not actually inductive and that it does not rely on an empirically *evidential* connection, since here again “the truth of the conclusion *is built into the meaning of its critical term*”—the critical term in this case being “someone who has learnt the things pertaining to the craft of *x*.” We can well imagine, for example, that Socrates would find it inconceivable that someone could *be*—and not just resemble—a genuine musician without learning, and so knowing, music.<sup>17</sup> We now need to ask whether (a) Vlastos is right to take *Ion* 540b–d and his analysis of it as representative of Socratic *epagôgê*, and whether (b) that passage’s pseudo-inference would count as the sort of *epagôgê* Aristotle thought particularly characteristic of Socratic philosophizing.<sup>18</sup>

This second question (b) is difficult to answer because of Aristotle’s notorious failure to provide an adequate account of *epagôgê* and its relation to what we would count as actual induction.<sup>19</sup> Nevertheless, consider Aristotle’s brief definition of *epagôgê* in the *Topics* (I.12)

. . . induction (*epagôgê*) is a passage (*ephodos*) from particulars to universals—e.g., the argument that supposing the skilled pilot is the most effective, and likewise the skilled charioteer, then in general that the skilled man is the best at his particular task. (*Top* 105a11–16)

The claim that this definition *does* apply to the kind of *epagôgê* Aristotle found characteristic of Socratic philosophizing is suggested by the resemblance of Aristotle’s example to the argument of *Euthydemus* 279d–280b. There Socrates maintains that since wise flute players have the best fortune/success at flute playing, wise grammarians at reading and writing, wise pilots at navigating the dangers of the sea, wise generals in producing victory in battle, and wise doctors in producing health, wisdom produces fortune/success in every case (see also, e.g., *Meno* 86e–89a). Here, however—and in answer to our questions (a) and (b) above—we are given reason to wonder whether Aristotle would find that Vlastos’ *Ion* passage provides a typical or paradigmatic instance of Socratic *epagôgê*. For, first of all, the conclusions of both Aristotle’s example and the argument of the *Euthydemus*—unlike the result of the non-inference of the *Ion*—do appear to admit of disconfirmation. It is at least conceivable, for example, that there should be a knowledge whose successful application is consistently stymied by associated psychological or other factors that



are immune to rectification; for example, knowledgeable gymnasts who are not as successful as those more ignorant practitioners who simply have an instinctive or practiced-and-acquired knack for such activity would be one such case (see, e.g., *La.* 185e–186a). Secondly, scholarly studies of Aristotelian *epagôgê* contend that he admits of three kinds of it, only one of which provides a model for the *Ion*'s example: (1) coming to know the universal for the first time without reflection by intuiting a single instance (e.g., suddenly realizing that all triangles have three interior angles after encountering merely one triangle); (2) coming to know the universal for the first time as a result of reflecting on or reasoning about one or more instances (e.g., concluding that all birds have hearts on the basis of a sample of various types of birds); and (3) knowing the universal already, and then as a result of reflecting on or reasoning about one or more instances, coming to recognize that those instances fall under that universal (e.g., knowing of some figure, and then realizing that the figure is a triangle).<sup>20</sup> Here kind (3) seems to come closest to capturing the non-inferential, explanatory epagogic passages of *Ion* 540b–d, *Protagoras* 332a–e, and *Gorgias* 460a–e, leaving kind (2) as the correct category under which to place Aristotle's example in *Topics* I.12 and the argument of *Euthydemus* 279d–280b. We may understand Aristotle, then, to have ascribed probabilistic inductive *epagôgê* to Socrates as well as explicative *epagôgê*.

In support of this, note how implausible it would be to hold that in the previous citation of the *epagôgê* at *Euthydemus* 279d–280b that Socrates is only trying to clarify what Cleinias already believes as opposed to leading him to hold a new belief. To wit: as part of the prelude to his *epagôgê*, Socrates asks Cleinias whether they have left anything out of their list of the things commonly taken to be the leading happiness-promoting goods—external, somatic, political, and characterological goods such as wealth, health, noble birth, conventional courage, and wisdom—and Cleinias replies “I don't think we are leaving any out” (*Euthd.* 279c4). And then, when Socrates prods him by proclaiming that “we are in danger of leaving out the *greatest* good of all,” and despite the fact that good fortune (*eutuchia*) was held by both popular and poetic opinion to be such a good, Cleinias still asks “which one is that?” (279c6). Even then, to achieve Cleinias' acceptance of good fortune as a good, Socrates has to place heavy rhetorical pressure on him by saying that “even quite *worthless* people” hold good fortune to be the greatest of the goods (279c7–8). Next, instead of *asking* Cleinias if he believes good fortune to be the same thing as wisdom, Socrates simply asserts that it is, and then claims that “even a *child* would know” that it is (279d6–7).<sup>21</sup> And no wonder Cleinias did not *already* believe this, we are told, for he is “so young and simple-minded” (279d7–8). This characterization is good evidence that Socrates does not view Cleinias to be a useful and reliable source of belief on the topic at hand, a source who merely needs to have his already-existing beliefs brought to light.<sup>22</sup>

In any case, if Vlastos and like-minded others are correct to cast *epagôgê* as simply a method of discovering what interlocutors already believe, then when Socrates purportedly wants to ascertain merely, say, whether Charmides believes that the swift and not the quiet are finest (*Chrm.* 159b–160d; see below), why does Socrates not just ask him? Or, granting that a few illustrations might be required to make clear

the nature of the question to Charmides, why does Socrates use *seven* examples, when clearly two would do for *explanatory* purposes? Moreover, many *epagôgê* are not prefaced with the sorts of linguistic markers that Socrates often uses to indicate that explanation is underway (see, e.g., *Euphr.* 10a5: “Then I shall try to put it (the question whether the pious is loved by the gods because it is pious, or pious because it is loved by the gods) more clearly”; cf. 10b10–c1 and 12a–d). Note, finally, *Statesman* 277c–278c, where we are told that there is a way of leading a person on—*epagein*, *anagein* (278a)—to the knowledge that he or she lacks. This way is illustrated by using the example of young children learning letters and then syllables—here these children are said to have “identified letters *correctly*. . . [so that] in the end, by this method, when the *rightly identified* letters have been shown to them and set alongside all the unknown letters. . . the teacher will achieve his aim.” This “method of example proceeds. . . when a factor identical with a factor in a less-known object is *rightly believed* to exist in some other *better-known* object in quite another sphere of life. . . [which] makes it possible for us to achieve a *single true judgment* about each of them. . .” (my italics). So, then, whether one is a fan of the strict, problematic interpretation of the Principle of the Priority of Definition or not (see n. 26), this single passage should lead us to the view that at least Plato, and then Aristotle, saw in Socrates’ tiresome talk of “cobblers, builders, and metal workers” (*Memorabilia* 1.2.37) not simply a method for discovering what interlocutors happen to believe, but the ancestor of those inductive inferences that lead interlocutors on to what they *ought* to believe. So, then, Vlastos—but not Santas or Aristotle—appears to have overlooked the considerable presence of probabilistic, inductive, persuasive *epagôgê* in our texts.

Let us now consider a few such arguments in order to further our understanding of Socratic *epagôgê* (a requirement of this paper, since, again, it appears that neither Socrates nor Plato ever offered a theory of *epagôgê*).

## 4.2 Socratic Epagogic Induction

As we saw above, we can find Socrates employing what seem to be epagogic arguments from a single proposition or set of coordinate propositions that serve as premises to another coordinate proposition—for example, “expert pilots have the best success at sea-faring; therefore, expert builders have the best success at building” (e.g., *Ap.* 27b; *H. Ma.* 284a–b)—that we may understand to be cases of inductive analogy. Santas himself offers an account of Socrates’ use of inductive analogy in a separate section prior to his examination of Socrates’ use of inductive generalization (Santas 1979, pp. 138–47). However, since Socrates does not himself seem to offer or to possess a theory of *epagôgê*, and because—as we have seen—Socrates sometimes employs non-inferential explanatory *epagôgê*, interpretive caution is in order when encountering any putative inductive analogy. For example, Santas interprets the *epagôgê* at *Laches* 184d–185a as follows (Santas 1979, p. 140):



- (P1) The consultation as to what exercises our sons should learn for a coming contest is a consultation about means to ends, and in it we must be guided by experts on those means and ends and not by majority (vote) of non-experts.
- (P2) The consultation as to whether our sons should learn fighting in armor is a consultation about means to ends.
- (C1) Therefore, in it (the consultation as to whether our sons should learn fighting in armor) we must be guided by experts and not by a majority (vote) of non-experts.

This argument does indeed have the form of an inductive analogy, but the text itself is more easily and naturally interpreted as holding that:

- (E1) The consultation as to what exercises our sons should learn for a coming contest is a consultation about means to ends, and in it we must be guided by experts on those means and ends and not by majority (vote) of non-experts.
- (P1) E1 introduces and illustrates the truth that “it is by knowledge that one ought to make [important] decisions, if one is make them well; not by majority rule” of non-experts (184e8–9).
- (P2) The consultation as to whether our sons should learn fighting in armor calls for an important decision.
- (C1) Therefore, in it (the consultation as to whether our sons should learn fighting in armor) we must be guided by experts and not by a majority (vote) of non-experts.

And this form is, of course, purely deductive, and not inductive at all.

Again, Santas finds the argument of *Crito* 47a–48e to be an inferential analogy (Santas 1979, pp. 145–6):

- (P1) In the matters of what actions promote the good condition of the body, health and strength, and what actions destroy the bad condition of the body, disease, (a) it takes experience and knowledge to determine which actions promote health and which destroy disease, i.e. which actions benefit and which harm the body; and therefore (b) if one wants to develop health and strength, the good condition of the body, and to avoid diseases, the bad conditions of the body, one should consult and follow the advice of those who have experience and knowledge of these matters and one should disregard the opinions of the public (the non-experts on these matters).
- (P2) In the matters of what actions promote the good condition of the psyche, virtue, and what actions destroy the bad conditions of the psyche, the vices, (a) it takes experience and knowledge to determine which actions promote virtue in the psyche and which destroy vice, i.e. which actions benefit the psyche and which harm it.

- (C1) Therefore, (b) if one wants to develop virtue, the good condition of the psyche, and avoid vices, the bad conditions of the psyche, one should consult and follow the advice of those who have experience and knowledge of these matters and one should disregard the opinions of the public (i.e., the non-experts on these matters).

This is an elegant reconstruction, and one that Santas rightly claims to exhibit argumentative strength (Santas 1979, p. 146). However, just as we need to beware of mistaking simple deductive inferences for inductive analogies, so we also need to beware of mistaking applications of inductive generalizations (“statistical” type inductions) for simple analogies. For in the case of *Crito* 47a–48e, we find a crucial phrase following the assertion that one who disobeys the expert concerning the body will incur bodily harm (47c1–7), namely:

And isn't the same true *in other cases*, dear Crito? No need to go through them all, but, in particular, cases of just and unjust things, shameful and fines one, good and bad ones—in cases of what we're now deliberating about . . . ? (*Cri.* 47c8–11)

This claim suggests that Socrates really means for us to understand his argument as holding that since in *all* the relevant cases of deliberation we care to imagine, one should consult and follow the advice of those who have the relevant experience and knowledge and disregard the opinions of the public, the same holds for decisions concerning the health of the psyche. This looks to be more an inductive generalization—or even a complete enumeration—to the general principle that in all matters of deliberation we should consult the expert and not the public, from which it follows *via* a statistical-type induction—or even deductively—that we should consult the expert concerning the health of the psyche. So, then, given this result, we are wise to be careful in our identification of inductive analogies in the Platonic corpus. But this result also bids us to make a close examination of a few cases of what appear to be actual Socratic inductive generalizations.

I think we find typical instances of probabilistic, inductive Socratic *epagôgê* in the arguments of (1) *Charmides* 159b–160d, (2) *Charmides* 167c–168b, and (3) *Memorabilia* 1.2.9 (see also, e.g., *Euphr.* 7a6–8a8; *La.* 192b9–193d8).

First, the *elenchos* of *Charmides* 159b–160d can be formalized as follows:

- (1) (P1) Temperance is some sort of quietness. (159b5–6)  
 (P2) Temperance is a fine thing. (159c1–2)  
 (P3) Writing swiftly and sharply, not quietly, is fine  
 (P4) Reading swiftly and sharply, not quietly, is fine.  
 (P5) Playing the lyre swiftly and sharply, not quietly, is fine.  
 (P6) Wrestling, boxing, fighting in the *pankration* swiftly and sharply, not quietly, is fine.  
 (P7) Running, and jumping swiftly and sharply, not quietly, is fine. (159c3–d3)  
 (C1) So, with respect to the body, not the quiet, but the swiftest and sharpest are finest. (From (P3–7); 159d4–7)

- (C2) So, with respect to the body, not quietness, but swiftness would be more temperate. (From (P2) and (C1); 159d10–12)
- (P8) Learning swiftly, vehemently, and sharply, not quietly, is fine.
- (P9) Teaching swiftly, vehemently, and sharply, not quietly, is fine.
- (P10) Recollecting swiftly, vehemently, and sharply, not quietly, is fine.
- (P11) Shrewdness involves swift, vehement, and sharp actions, not quiet ones, and thus is fine.
- (P12) Understanding at the writing teacher’s and lyre teacher’s desks, etc., swiftly, vehemently, and sharply, not quietly, is fine.
- (P13) Inquiring, and taking council swiftly, vehemently, and sharply, not quietly, is fine. (159e1–160b2)
- (C3) So, in every case, both concerning the mind and the body, the quick and most sharp things seem to be finer than the most slow and most quiet things. (From (C2) and (P8–13); 160b3–6)
- (C4) “Then temperance won’t be a sort of quietness. . . at least according to this argument (*ek ge toutou tou logou*), since being temperate it must be fine.” (From (C2) and (C3); 160b7–9)<sup>23</sup>

Since we are concerned here with Socratic *epagôgê* and not Socratic *elenchoi* in general, I will simply bypass the several logical problems posed by this argument (on which see Santas 1973) by following Santas’ analysis—making the reasonable assumptions that the opposition “quickly versus quietly” is sensible on the grounds that Socrates presupposes that “in the majority of cases when people do things quietly they also tend to do them slowly” (Santas 1973, p. 115, n. 9) and that temperance and the other virtues are the only truly praiseworthy things (Santas 1973, p. 117). What I wish to draw attention to instead is that Socrates summarizes the results of this argument by saying that

. . . temperance is not quietness. . . . And of two things one is true: either never, *or in very few cases (oligachou; 160c1)* do the quiet actions in life appear to be better than the quick and energetic ones, or supposing that of the nobler actions there are as many quiet ones as quick and vehement ones; still, even if we grant this, temperance will not be acting quietly any more than acting quickly and energetically, either in walking or talking or in anything else. (*Chrm.* 160b7–d3; my italics)

The italicized phrase of this passage indicates that Socrates takes his (C1) and (C3) to be probabilistic, and thus, that the epagogic argumentation of (P1–C4) cannot fall under the Vlastosian model.<sup>24</sup> The reasoning here appears, rather, to employ instances of Aristotelian type 2 *epagôgê* (coming to know the universal for the first time as a result of reasoning about one or more instances). However, this mode of inference also seems to be a case of intuiting or generating a universal on the basis of an insufficiently large sample; thus, it may suggest that Socrates’ probabilistic *epagôgê* is a charming—but eccentric—Socratic practice, like walking on ice in your bare feet (*Smp.* 220b–c). What defense of this procedure, then, can we offer?

First, we should avoid anachronism by realizing that pre-Aristotelian, pre-Stoic inductive argumentation generally displays a startling indifference to considerations

of sample size or bias. We should therefore give credit to Socrates for employing, typically, more than two items of evidence in his epagogic reasoning. Next, a moment's reflection on our own everyday uses of inductive generalization should also cause us to moderate, if not reconsider, our criticism. Consider, for example, the sample size we deem sufficient when we are already convinced of the relative uniformity of our evidence, as in the case of our everyday curiosity as to the current gasoline prices within a relatively uniform region (say, Franklin County, Maine). Here we would be right to think that we have accounted sufficiently for variations in the wholesale prices offered to dealers, the market effects of local demand, distance from distribution centers, discount filling stations, and so on, if we take note of the gasoline prices at merely six scattered filling stations (including a discount station): that number would be adequate (for practical purposes) for us to cogently affirm the universal claim that, say, "Gasoline prices are over \$2.00 a gallon in Franklin County." Of course, here we are warranted in thinking that our conclusion is highly probable because we have antecedent assurance that all six stations are really stations, that wholesalers' prices are usually within a few cents of one another, that dealers adjust their prices in competition with other dealers, and so on. It seems to me, then, that we ought to investigate whether Socrates reasons epagogically on the basis of analogous presuppositions.

To this end, consider the classic account of the relation between epagogic instances and definition found in the *Euthyphro* (5c–6e). Here we learn that what Socrates hopes to elicit from Euthyphro is an account of the *one eidos* of piety: that unique, self-same universal characteristic by which *all* pious acts are made pious, the characteristic that Euthyphro had earlier agreed was the object of their search; it is "that character in respect of which" such things as the virtues "do not differ at all, but are all the same. . . [for e]ven if they are many and various, yet at least they all have some common character (*eidos*) which makes them virtues" (*Meno* 72c–d; cf. *La.* 191d). Such an *eidos* is a *pattern* (*paradeigma*), and so possessing a complete account of piety's *eidos* would put its possessor in a position that (together with other factors) would allow him or her to recognize reliably actions that are pious to be pious, and to distinguish them from those that are not. Here it is clear that universal characteristics such as piety are taken by Socrates to be objective items with real natures *by which* the things possessing them are correctly named by using general terms such as "piety" (e.g., *Euphr.* 5c–6e, *Prt.* 330c, *H. Ma.* 287c–d; cf. Aristotle, *Met* A.6 987b1–5: "[Socrates] seeking the universal in these ethical matters. . . fixed thought for the first time on definitions").<sup>25</sup> Moreover, Socrates' elenctic method of testing definitional candidates by subjecting them to counterexamples appears to assume that virtually every human being believes that he or she is frequently able to accurately identify particular instances of some universal.<sup>26</sup> This suggests, then, that Socratic *epagôgê* of Aristotelian type 2 (coming to know the universal for the first time as a result of reasoning about one or more instances) presuppose that the few cases they employ are sufficiently representative and pure—like the gasoline stations of our example—for the interlocutor to affirm the presence of an embedded, defining universal in each case, and thus, that universal's associated characteristics so as to warrant and lead an interlocutor to affirm a more general and

probable conclusion.<sup>27</sup> This means, of course, that contrary to Robinson's account of *epagôgê* as (A) Intuiting the Universal where this yields *certainty*, that Socrates saw such intuitions of a universal as standing in need of corroboration through a sampling of cases and as only providing probable results. To buttress this suggestion, consider the epagogic reasoning deployed at both (2) *Charmides* 167c–168b and (3) *Memorabilia* 1.2.9 (cf. 4.2.6–7). The argument from *Chrm.* 167c–168b:

- (2) (P1) There is no vision of vision.
- (P2) There is no hearing of hearing.
- (P3) There is no sense of itself and the other senses.
- (P4) There is no desire of desires.
- (P5) There is no wish that wishes only for itself and other wishes.
- (P6) There is no love that loves only itself and other loves.
- (P7) There is no fear that fears only itself or other fears.
- (P8) There is no opinion that is an opinion only of itself and of other opinions.
- (C1) So, there is no mental state that is only of itself (implicit).
- (C2) So, there is no science of itself and of the other sciences (“It is certainly a curiosity if it really exists. We must not however as yet absolutely deny the possibility of such a science, but continue to inquire whether it exists” [168a10–11; cf. 168b-169c]).

Finally, the argument from *Memorabilia* 1.2.9:

- (3) (P1) We rightly attempt to choose the best pilot not by lot but on the basis of his or her knowledge of piloting.
- (P2) We rightly attempt to choose the best builder not by lot but on the basis of his or her knowledge of building.
- (P3) We rightly attempt to choose the best flute player not by lot but on the basis of his or her knowledge of flute playing.
- (C1) We rightly attempt to choose the best craftsman (to accomplish a task of importance) not by lot but on the basis of his or her knowledge of the relevant craft.
- (P4) Governing a state is a task of importance and there is a craft of it (statesmanship).
- (C2) We ought not to choose a statesman by lot but through a test of a candidate's knowledge of statesmanship.<sup>28</sup>

Both (2) and (3) are representative of type 2 *epagôgê*; in particular, (2)'s (C2) makes it clear that the conclusion is thought to follow probably, for the most part, and not necessarily.<sup>29</sup> Argument (3)'s (C1) would also seem probabilistic, since one can imagine a counterexample to its claim—namely, the existence of a craft whose expertise is sufficiently difficult to test for (say, the craft of divination) that it is better to leave the choice of its best practitioner to the lot (i.e., to divination).<sup>30</sup>

Although it is true that in this case—unlike the *epagôgê* of (2)—Socrates does not explicitly claim that (C1) is only probably, generally true, it appears that Socrates often leaves his modal qualifiers implicit. For, again, since Socrates' chief aim is to elicit agreement to a general claim from an interlocutor in the course of his elenctic testing of that interlocutor, we may expect Socrates to deem such specifications to be of distinctly secondary importance. Both arguments, then, strike me as representing instances of inductive reasoning, instances that are reasonably successful on both Socratic and modern criteria. Argument (2) from (P1) to (C1), in particular, targets an appropriately diverse sample of mental states, discerning in each a common characteristic, namely, lack of reflexivity.

The preceding arguments, then, and many others in the Socratic *testimonia*, validate the 1979 insight of Santas that marked an important correction to the 1953 work of Robinson on *epagôgê*. That is, they give us good reason for thinking that when need be, Socratic *epagôgê* could rise from the level of explanation (one sort of *epagôgê*—the Vlastosian explicative *epagôgê*) to the level of respectable Socratic induction—conceived here as Probable Inductive Generalization Employing a Survey of Coordinate Cases involving Intuition of the Universal (but thus not yielding certainty, as in conception [A] of Intuition of the Universal). If so, then we should hope to see future investigations of Socratic *elenchoi* that pay more detailed attention to this pivotal and too-long-neglected facet of Socratic argumentation. Secondly, I have indicated how advocates of anti-constructivist interpretations of the *elenchos* need to provide a better account of the place of Socratic *epagôgê* in what they allege to be Socrates' single-minded use of the *elenchos* to pursue knowledge-testing of knowledge-professing interlocutors.<sup>31</sup>

## Notes

1. It was Santas' book that prompted my initial investigations into the topic of *epagôgê*, which then resulted in McPherran 2007, from which this paper derives.
2. According to Eustathius (*In Iliadem* 4.736.8–11), it was Homer (at *Iliad* 23.313–25) who first used “the induction [*epagôgê*] of the philosophers, namely, the argument establishing the universal from particulars,” Ausland (2002, p. 49). Socrates is, of course, commonly portrayed as having conducted his philosophical mission primarily by examining interlocutors using the technique we call “the *elenchos*.” Consequently, it is natural that the structure and purposes of those examinations are of persistent scholarly interest. But we also have reason to believe that Socrates was preceded by other thinkers in cultivating the systematic use of this procedure (although it was he who first put it to the revolutionary, virtue-promoting ends of philosophical self-examination of the *psychê*). It is, for example, commonly thought that Socrates derives his interrogative method from the rhetorical techniques of eristic and *antilogikê* employed by the Sophists; see, e.g., Guthrie (1971b, pp. 41–4); and Kerferd (1981, ch. 6); see also Leshner (2002) and Ausland (2002). That explains why it is that when Aristotle assesses Socrates' contributions to the development of philosophical method he does not mention the *elenchos per se* (Robinson 1953, pp. 46–8). Scott (2002) is a recent collection of essays on the *elenchos*. Let me note here that this paper does not attempt to identify the views of the historical Socrates, but rather, those of the cross-dialogue, literary figure that emerges from the Socratic dialogues of Plato in concert with the recollections of Xenophon and others (e.g., Aristotle). These portraits constitute a mosaic of the characteristics, methods,

views, and activities of a Socrates who manifests distinctly different philosophical attitudes from those expressed by the Socrates of Plato's *Republic* and other such constructive and, arguably, later dialogues. This qualification permits me to avoid the difficult issue of how we might accurately arrive at the views and activities of the actual teacher of Plato, yet still allows us to confront many of the most interesting questions Plato's works provoke. There is not sufficient space here to address the complex issue of whether and how we might legitimately use the testimony of Aristotle in conjunction with that of Plato's dialogues and Xenophon's work to triangulate to the views of the historical Socrates in the manner of Gregory Vlastos (1991, chs. 2, 3) (but see, e.g., McPherran [1996, ch. 1.2]).

3. See also Gulley (1968, pp. 13–21); and Guthrie (1971a, pp. 105–10), who offers a brief overview of *epagôgê* that is essentially a précis of Robinson (Ausland 2002 offers a useful overview that traces the historical antecedents of Socratic *epagôgê*).
4. Over the course of his many outlines of Socratic *elenchoi*, for example, Benson (2000), only notes one case of epagogic/inductive reasoning in passing (p. 48 n. 55). Again, Dancy (2004, pp. 138–41), finds the epagogic list of cases at *Euthyphro* 10a–d to inductively support Socrates' general claim at 10b11–c5, but the text (at 10a5 and 10b10–c1) makes it quite clear that Socrates is not using his list of cases to *argue* for the truth of a claim, but, rather, to *explain* his claim to Euthyphro.
5. Robinson (1953, p. 207); cf. *An. Pr* II.24; *Rhet* 1393a26–b3 on “argument by example (*paradeigma*).”
6. *Ibid.*, p. 40. Robinson understands *epagôgê* to be part of a larger phenomenon that he calls “the use of cases,” a technique that ranges from illustrative example, to analogical inference from a single case, to complete enumeration (pp. 42–5).
7. *Ibid.*, p. 37. However, Robinson implicitly concedes that a brief epagogic survey of instances may yield a probability when he states that the review of instances displayed at *Charmides* 167–9 makes it merely “very doubtful” that there could be knowledge of knowledge (p. 39).
8. For example, *pithanos* (*Phd.* 88d2); *pithanologia* (*Tht.* 162e9); see also Socrates' awareness of the need for samples at, e.g., *La.* 185e–186a. Socrates would also have been familiar with the Sophists' penchant for arguments from probability, a trait reputedly fostered by Corax's account of forensic arguments from probability in his *The Technique of Speaking*; see Freeman (1963, p. 31).
9. Benson (2000, p. 139), appears to hold that the issue of whether an interlocutor's belief is well-grounded is *never* a concern in any elenctic encounter, but that stance provokes the question as to why Socrates bothers to employ as many epagogic arguments, containing as many examples as he does (on which, see below). On Socrates' typical demand that interlocutors “say what they believe,” see, e.g., Vlastos (1994) and Beversluis (2000, ch. 2).
10. Some Socratic propositions, however, do seem to be foundational and certain for Socrates—for example, the claim “the gods are wise” (which he may have thought to be true by definition)—and hence, we never observe Socrates subjecting statements of this sort to elenctic examination.
11. Thus, as Annas (1976, p. 154), points out, *epagôgê* “is not . . . the same as modern induction.” Similarly, Vlastos (1991, p. 267) rightly complains that “the mistranslation of Aristotle's *epaktikoi logoi* as ‘inductive arguments’ is virtually ubiquitous in the scholarly literature.”
12. Thus, the term “intuitive induction” does not designate what I earlier termed “Intuition of the Universal” (A).
13. Rather, induction would show up in Socratic philosophizing only insofar as Socrates might take the dialogical resilience of his moral beliefs (e.g., that it is better to suffer an injustice than to do one) to constitute probabilistic warrant of the truth of those beliefs. On this issue, see Vlastos (1994, pp. 27–8) and Brickhouse and Smith (2000, p. 89).
14. This would be a somewhat odd result for Vlastos, given his status as the originator of the constructivist interpretation of the Socratic *elenchos* (see Vlastos, “The Socratic Elenchus”): that is, again, the view that the *elenchos* as a practice of eliciting contradictions from an interlocutor is used constructively to test not only the consistency of an interlocutor's



- belief set in order to adjudicate that interlocutor's claim to knowledge, but to also secure Socrates' moral and action-guiding tenets. Recent expressions of opposition to this idea—the “anti-constructive” position—include Benson (2000, chs. 2–4) and Stokes (1986, pp. 1–35, 440–43).
15. A distinction complicated in part by the possibility that Socrates holds there to be a “true self” underlying each of our “false selves,” such that we may unconsciously harbor beliefs that are the negations of what we consciously believe and say we believe (which then explains, e.g., how “Callicles might not agree with Callicles” [*Grg.* 495d–e]); see Brickhouse and Smith (1994, ch. 3, sect. 6.1) on the idea of the true self in Socratic philosophy. This facet of Socratic psychology may foreshadow Plato's later elaborated view that learning is recollection. On the relation between Aristotelian *epagôgê* and the Platonic doctrine of Recollection, see Gifford (1999), and now LaBarge (2004).
  16. See McPherran (2004, pp. 11–33), for an account of how Socrates uses the *elenchos* to diagnose his “patients,” namely, interlocutor-patients such as Charmides and Critias.
  17. See Santas' analysis (1979, pp. 148–50) of the *epagôgê* at *Hippias Minor* 367b–369a for the view that “those who have the greatest power to tell intentional lies are those who have the greatest power to tell truths” is similarly flawed. Santas claims that the passage is “a perfect example of an inductive generalization” of the form P1 “a, b, c. . . have S and P”; P2 “Nothing has been observed to be S and not P”; C1 “Therefore, probably all S are P” (p. 149). However, the ability to tell lies (reliably) in a subject matter M is logically related to the ability to discern truths from falsehoods within the scope of M. Hence, it is not merely probable that those who can reliably discern truth in M can also reliably discern falsehood in M.
  18. By taking the *Ion* passage as a paradigmatic *epagôgê*, Vlastos abandoned his earlier view, according to which Socrates *does* employ arguments that appear to be *bone fide* inductions (i.e., probable inferences from a series of cases) but “fails to recognize the importance” of the issue of whether the cited cases have sufficient homogeneity to warrant his generalizations (Vlastos 1956, p. xxxvii n. 45); cf. Gulley (1968, p. 21); and Vlastos (1994, p. 4 n. 15).
  19. Smith (1995, p. 33). An “actual induction” is any non-deductive argument whose conclusion purports to be made more likely true than not by the truth of its premises.
  20. See McKirahan (1983, p. 9); also Leshner (1973), Hamlyn (1976), and now LaBarge (2004). Aristotle does appear to connect what we would identify as induction with *epagôgê* at *An. Post* II.19 100a15–100b5.
  21. Cf. *Smp.* 204b, where the claim that “a child (*pais*) could tell you” is used to characterize something thought to be obviously true. That Socrates intends to make an assertion that will puzzle Cleinias, provoking him to unpack it, is something Plato communicates to us when he has Socrates report that Cleinias expressed *wonder* (*thauma*, 279d8) at his claim.
  22. For a recent full discussion of this section of the *Euthydemus*' first protreptic, see McPherran (2005).
  23. Cf. the reconstructions and discussions by Benson (2000, pp. 71–3); and Dancy (2004, pp. 93–8).
  24. Note that Dancy (2004, p. 93), terms the “long list of cases” Socrates produces between *Charmides* 159c3 and 160b6 “a Socratic induction.” Dancy (2004, p. 108) also finds induction at work at *Charmides* 165c–d.
  25. That this principle is a quite general one, and is not restricted to virtue concepts, is indicated by the *Lysis* at 217c–d, where white hair is said to be of the same sort as (*hoion*) what is present to it.
  26. This claim, however, raises the issue of whether Socrates is committed to a version of the Principle of the Priority of Definition (PD) that would convict him of committing “the Socratic Fallacy” (e.g., holding that [PD] to know that some purported case of F is an instance of F-ness one must have definitional knowledge of what F-ness is, but then also disclaiming knowledge of what F-ness is, while yet claiming to know that some particular Fs are instances of F-ness). Brickhouse and Smith (2000, ch. 3, sect. 2) argue that Socrates is not committed to a problematic version of PD; but Benson (2000, ch. 6, esp. pp. 131–41) contends that Socrates



- is so committed, but that elenctic examinations do not require either his interlocutor or himself to claim to *know* that any purported instances of F-ness are *in fact* F. Rather, Benson holds that since the goal of an *elenchos* is to detect whether an interlocutor has knowledge of what F-ness is, and since the revelation that an interlocutor holds contradictory beliefs concerning F-ness is sufficient to establish that he or she lacks such knowledge, all that any successful *elenchos* requires is that the interlocutor reveal that he or she does *believe* that some item is an instance of F-ness. However, again, Benson does not explain why Socrates nevertheless appears to use *epagôgê* not only to *clarify* some claim in order to elicit an interlocutor's beliefs on the topic at hand, but also to *argue for* some claim so as to produce warranted belief. I also think it odd that a Socrates who was committed to a problematic version of PD would not, as he leads his interlocutors through the steps of an *epagôgê*, stop them after they agree that, say, the expert in medicine is a doctor (*Grg.* 460b), and then ask them "But *why* do you believe that?"
27. Although for Aristotle induction is less a method of justification than it is a causal process, Socrates leaves it quite unclear as to how we are to understand the mechanisms underlying his own use of type 2 epagogic reasoning. However, the idea that Socrates is trying to causally implant beliefs rather than to justify them would provide a solution to the Number of Instances Problem: we can suppose that Socrates simply uses whatever number of cases he perceives likely to bring his particular interlocutor to assent to the universal claim he, Socrates, is driving at (so as to affirm a claim the interlocutor did not believe before). On the other hand, this solution makes Socrates look like more of a dogmatist than some might like, and it raises the question of what Socrates takes to warrant his own (epagogically-derived?) moral beliefs.
  28. Cf. Aristotle:

The illustrative parallel is the sort of argument Socrates used: e.g., "Public officials ought not to be selected by lot. That is like using the lot to select athletes, instead of choosing those who are fit for the contest; or using the lot to select a steersman from among a ship's crew, as if we ought to take the man on whom the lot falls, and not the man who knows most about it." (*Rhetoric* II.20 1393b4–8)

29. As it should, in view of the pleonexic phenomenon of desiring to have greater desires (as observed at, e.g., *Grg.* 491e–494d), which would falsify both (2)'s (P4) and (C1). Note that Dancy (2004, pp. 102–3) finds the passage from 167b to 169c to be "a Socratic induction aiming at establishing the generalization, stated in 169a3–4: (G) None of the things that are of a nature to have its faculty (*dunamis*) relative to itself (*pros heauto*)."
30. Athenians commonly held that the lot made it possible for the knowledgeable gods, rather than fallible humans, to decide a matter at issue; Sourvinou-Inwood (1990, p. 321). Arguably, Socrates believes that "... what the gods have granted us to do by help of learning, we must learn; what is hidden from mortals we should try to find out from the gods by divination. . ." (*Memorabilia* 1.1.9; Marchant transl.); see McPherran (1996, ch. 4), on Socrates' acceptance of a rationally-revised form of divination.
31. My thanks to Hugh Benson, Daniel Cohen, Jan Kaufman, and Scott LaBarge for their comments on an earlier version of this paper. I am also indebted to Georgios Anagnostopoulos for the invitation to contribute to this volume and to participate in the Santas fest held in Pyrgos Greece, August, 2006. Thanks to my audience on that occasion, and most all, to Jerry Santas for his many important contributions to my education and professional life over the years.

## Bibliography

- Annas, J. 1976. *Aristotle's Metaphysics: Books M and N*. Oxford: Oxford University Press.  
 Ausland, H. 2002. "Forensic Characteristics of Socratic Argumentation." In edited by G.A. Scott.  
 Benson, H. 2000. *Socratic Wisdom*. Oxford: Oxford University Press.

- Beverluis, J. 2000. *Cross-Examining Socrates. A Defense of the Interlocutors in Plato's Early Dialogues*. Cambridge: Cambridge University Press.
- Brickhouse, T.C., and N.D. Smith. 1994. *Plato's Socrates*. Oxford: Oxford University Press.
- Brickhouse, T.C., and N.D. Smith. 2000. *The Philosophy of Socrates*. Boulder, CO: Westview Press.
- Dancy, R.M. 2004. *Plato's Introduction of Forms*. Cambridge: Cambridge University Press.
- Freeman, K. 1963. *The Murder of Herodes, and Other Trials from the Athenian Law Courts*. New York: Norton.
- Gifford, M. 1999. "Aristotle on Platonic Recollection and the Paradox of Knowing Universals: *Prior Analytics* B.21 67a8–30." *Phronesis* 44:1–29.
- Graham, D. 1990. "Socrates, the Craft Analogy, and Science." *Apeiron* 23:1–24.
- Gulley, N. 1968. *The Philosophy of Socrates*. London: Macmillan.
- Guthrie, W.K.C. 1971a. *Socrates*. Cambridge: Cambridge University Press.
- Guthrie, W.K.C. 1971b. *The Sophists*. Cambridge: Cambridge University Press.
- Hamlyn, D.W. 1976. "Aristotelian *Epagôgê*." *Phronesis* 21:167–84.
- Kerferd, G.B. 1981. *The Sophistic Movement*. Cambridge: Cambridge University Press.
- Kraut, R. 1983. "Comments on Gregory Vlastos, 'The Socratic Elenchus'." *Oxford Studies in Ancient Philosophy* 1:59–70.
- LaBarge, S. 2004. "Aristotle on 'Simultaneous Learning' in *A. Post* I.1 and *An. Pr* II.21." *Oxford Studies in Ancient Philosophy* 27:117–215.
- Leshner, J. 1973. "The Meaning of *Nous* in the *Posterior Analytics*." *Phronesis* 18:44–68.
- Leshner, J. 2002. "Parmenidean *Elenchos*." In edited by G.A. Scott.
- McKirahan, R.D. 1983. "Aristotelian *Epagoge* in *Prior Analytics* 2.21 and *Posterior Analytics* 1.1." *Journal of the History of Philosophy* 21:1–14.
- McPherran, M.L. 1996/1999. *The Religion of Socrates*. University Park, PA: Pennsylvania State University Press.
- McPherran, M.L. 2001. Review of John Beverluis' *Cross-Examining Socrates. A Defense of the Interlocutors in Plato's Early Dialogues*. *Journal of the History of Philosophy* 38:583–84.
- McPherran, M.L. 2002/2005. "Justice and Pollution in the *Euthyphro*." *Apeiron* 35:1–25; Reprinted in Kamtekar, R., ed. *Critical Essays on Plato's Euthyphro, Apology, and Crito*, 1–22. Lanham, MD: Rowman & Littlefield.
- McPherran, M.L. 2003. "The Aporetic Interlude and Fifth *Elenchos* of Plato's *Euthyphro*." *Oxford Studies in Ancient Philosophy* 25:1–37.
- McPherran, M.L. 2004. "Socrates and Zalmoxis on Drugs, Charms, and Purification." *Apeiron* 37(1):11–33.
- McPherran, M.L. 2005. "'What Even a Child Would Know': Socrates, Luck, and Providence at *Euthydemus* 277d–282e." *Ancient Philosophy* 25(1):1–15.
- McPherran, M.L. 2007. "Socratic *Epagôgê* and Socratic Induction." *Journal of the History of Philosophy* 45.3: 347–64.
- Robinson, R. 1953. *Plato's Earlier Dialectic*. Oxford: Oxford University Press.
- Santas, G.X. 1973. "Socrates at Work on Virtue and Knowledge in Plato's *Charmides*." In *Exegesis and Argument, Studies in Greek Philosophy Presented to Gregory Vlastos*, edited by E.N. Lee, A.P.D. Mourelatos, and R.M. Rorty, 105–32. Assen: Van Gorcum and Co.
- Santas, G.X. 1979. *Socrates*. London: Routledge.
- Scott, G.A., ed. 2002. *Does Socrates Have a Method? Rethinking the Elenchus in Plato's Dialogues and Beyond*. University Park, PA: Pennsylvania State University Press.
- Smith, R. 1995. "Logic." In *The Cambridge Companion to Aristotle*, edited by J. Barnes. ch. 2. Cambridge: Cambridge University Press.
- Sourvinou-Inwood, C. 1990. "What is Polis Religion?" In *The Greek City*, edited by O. Murray and S.R.F. Price. Oxford: Oxford University Press.
- Stokes, M.C. 1986. *Plato's Socratic Conversations: Drama and Dialectic in Three Dialogues*. Baltimore, MD: Johns Hopkins University Press.
- Taylor, A.E. 1932. *Socrates*. New York: D. Appleton.

- Vlastos, G. 1956. "Introduction." In *Plato's Protagoras*, edited by G. Vlastos. Upper Saddle River, NJ: Prentice-Hall.
- Vlastos, G. 1991. *Socrates, Ironist and Moral Philosopher*. Ithaca: Cornell University Press.
- Vlastos, G. 1994. "The Socratic Elenchus: Method is All." In *Socratic Studies*, edited by M. Burnyeat. Cambridge: Cambridge University Press.

# Chapter 5

## Socrates *Mythologikos*

Fred D. Miller, Jr.

### 5.1 Introduction: Why Did Socrates Turn to Mythmaking?

We owe a great debt to Gerasimos Santas, who has done so much to establish Socrates as a precursor to the modern analytic philosopher. In the *Apology*, as Santas notes, Socrates “conceives of himself as being commanded by god, or as having been stationed by god, with orders to philosophize.”<sup>1</sup> Santas’ focus is on the Socrates who questions himself and others, refutes their answers, offers arguments and criticizes them, and challenges ordinary beliefs by defending paradoxical doctrines. This portrait of Socrates the philosopher dominates Plato’s early dialogues.

Yet Socrates took a surprising turn at the end of his life, according to Plato’s *Phaedo*. When Cebes asked Socrates why he began to write poetry after he went to prison, putting the fables (*logoi*) of Aesop into verse and composing a hymn to Apollo, Socrates replied that he was responding to a recurring dream that commanded him to practice and cultivate the arts. Formerly he had tried to obey the dream by practicing the art of philosophy, but, after his trial, he thought that perhaps it was bidding him to compose poetry.<sup>2</sup> “I realized that a poet, if he is to be a poet, must make myths (*muthoi*) not arguments (*logoi*). Not being skilled at telling myths (*muthologikos*) myself, I took the myths (*muthoi*) I knew and had at hand, those of Aesop, and I versified the first ones I came across” (61b3–7). Socrates is better suited for this task now: “It is perhaps most appropriate for one who is about to depart yonder to examine and tell myths (*muthologeîn*) about what we believe that journey to be like” (61d10–e3).

Socrates’ conversion to poetry is puzzling in a number of ways. Shortly after denying that he is skilled at telling myths (*muthologikos*), he says that it is now appropriate for him to tell them (*muthologeîn*) because he is approaching death. He adds however that he is only recounting what he has heard from others, who are presumably more skilled at composing the myths in the first place. Why did his

---

F.D. Miller (✉)  
Bowling Green State University, Bowling Green, OH, USA  
e-mail: fmiller@bgnet.bgsu.edu

dream command him to tell myths if he is not skilled at doing so? And why does he deny that he is *muthologikos* when he seems so good at making up stories? Is this an instance of his notorious irony?<sup>3</sup> Socrates implies that making myths is the poet's job, while making arguments is the philosopher's. Yet, in turning from *logos* to *muthos* while in prison, he has not forsaken philosophy altogether, since he proceeds to argue at length for the immortality of the soul. But, at the end of the *Phaedo*, he turns from argument back to myth. What does all this suggest about Socrates' view of the relation of poetry to philosophy and about the relation of *muthos* to *logos*? Is the telling of myths a way of philosophizing, or is it merely incidental to his philosophical mission? If the latter, then it seems that the philosopher can achieve knowledge by means of arguments alone. Myths may be useful for persuading non-philosophers unable, unready, or unwilling to follow rational arguments. But it is hard to reconcile such a deflationary interpretation of myth-telling with Socrates' deathbed conversion to poetry and with his serious presentation of myths about the afterlife in the *Phaedo* and other dialogues. Catalin Partenie (2004, p. xxiii) states the problem in stark terms:

There is in Plato an inconsistency between what he says and what he does in his dialogues: on the one hand, he opposes myth to philosophical argument; but, on the other, he uses myths (and other fictional narratives) abundantly, and envelops his own philosophy in fictional narrative dialogues, in what seems a schizophrenic act of sabotage.

An opposed interpretation is that certain facts such as the ultimate destiny of the soul are unknowable by reason and can be represented only through myth. But this suggests that philosophy ultimately fails, that Plato's Socrates, like some modern romantics, forsakes dialectical reasoning for mystical feeling when he comes to the hardest questions.<sup>4</sup> But perhaps these two opposed interpretations—rationalism versus mysterianism—present a false dichotomy. Might Socrates' eleventh-hour turn to poetry suggest that the philosopher can acknowledge an indispensable function for myths without denying the efficacy of reason?<sup>5</sup> In any case, we have good reason to consider why Plato's Socrates regards myths about the afterlife as valuable.

This essay will consider Socrates' view of myth in the *Gorgias*, *Phaedo*, *Republic*, and *Phaedrus*.<sup>6</sup> Many scholars believe that these dialogues were composed in that order during Plato's middle period, although the dating of Plato's dialogues remains controversial. In any case I shall argue that, if the dialogues are considered in this order, Plato's views concerning myth and philosophy, as expressed by Socrates in these middle-period dialogues, will come into increasingly sharper focus.

## 5.2 “*Mythos*” and “*Logos*”

Let us begin with a brief discussion of Plato's terminology. “*Muthos*” may refer to a story or fable, although Plato often applies it to a story involving the gods or about the afterlife (in which sense it will be translated as “myth”).<sup>7</sup> “*Logos*” in the

broad sense refers to whatever is spoken or written, in which sense a *muthos* is also a *logos*.

The Greek word “*muthos*” is used to describe each of Plato’s four major myths concerning the afterlife: *Gorgias* 523a2, 527a5 (with qualification); *Phaedo* 110b1 and 4, 114d7; *Phaedrus* 253c7; *Republic* X 621b8. In all but the *Gorgias* the myth follows upon an argument (*logos*) that the soul is immortal. It seems significant that early on in the *Phaedo* (61b4, quoted above) Socrates draws a sharp distinction between *muthos* and *logos*: it is the job of the poet to make *muthoi* and that of the philosopher to make *logoi*. When the terms are contrasted this way, “*logos*” is translated as “argument,” “account,” “theory,” and so forth, and “*muthos*” as “story,” “tale,” or (in a religious context) “myth.”<sup>8</sup>

Plato’s Socrates does not always distinguish *muthos* from *logos* this sharply, however. In the *Republic* (II 376d9–10) he compares the discussion of something in *logos* to making up *muthoi*. His description of the ideal constitution is also said to combine *muthos* and *logos*: *muthologoumen logô(i)* (VI 501e4). Sometimes “*muthos*” and “*logos*” are used as if they were equivalent, for example, when Socrates mentions stories told to children (II 376e9, 377a4).<sup>9</sup> It is not Plato’s way to use words in a fixed and invariable manner, even within the same dialogue.<sup>10</sup> So we should not infer from the mere fact that Plato sometimes combines “*muthos*” and “*logos*” or uses the terms interchangeably that he is intentionally “blurring” the distinction between myth and reason. His view of the relation of myth to philosophy must be gleaned from a careful study of the context in which Socrates discusses these notions.

Finally, before proceeding, it is necessary to distinguish three main types of myths discussed by Plato.<sup>11</sup> Traditional myths, many of which come from poets like Homer and Hesiod, are filled with falsehood and corrupt those who hear them. Educational myths, such as the myth of the metals in the *Republic*, are employed by legislators and educators to instill virtuous beliefs and habits in children and benighted citizens. Philosophical myths are used by philosophers to augment (in some manner) rational arguments. The latter fall into four subtypes: historical (for example, the myth of Atlantis in the *Critias*), genetic (for example, the Demiurge stories of the *Timaeus* and *Statesman*), metaphysical (for example, the Sun, Line, and Cave in the *Republic*), and eschatological (the myths of the *Gorgias*, *Phaedo*, *Republic*, and *Phaedrus*). This essay is concerned with the last subtype: philosophical myths about the afterlife.

### 5.3 Myth in the *Gorgias*

The final myth of the *Gorgias* is distinctive in two ways: it does not follow a proof that the soul is immortal, and Socrates denies that it is a *muthos*. He instructs Callicles that a rational person fears injustice more than death. “For a soul to arrive in Hades filled with many injustices is the ultimate of all evils. And if you wish, I am ready to tell you an account (*logos*) [showing] that this is the case” (522e3–6).

Socrates continues, “Listen then, as they say, to a very fine account, which you, I think, will believe is a myth (*muthos*), but which I believe is as an account (*logos*); for what I am about to tell you I shall tell you as (*hōs*) the truth” (523a1–3). This way of putting it suggests that Socrates believes that what he will say is true, but he is not asserting it with complete confidence.<sup>12</sup> He proceeds to narrate a myth, attributed to Homer (523a3), concerning postmortem punishment and reward. He then affirms, “These are the things, Callicles, which I have heard and which I believe are true” (524a8–b1). He then offers an interpretation: death is the separation of the soul from the body, both of which retain in death characteristics possessed in life. The fate of the soul is determined by its inherent characteristics. Philosophers will be rewarded with a trip to the Isles of the Blessed, but there is no mention of reincarnation. Socrates concludes that he has been persuaded by these accounts (*logoi*) although they may look like mere myths (*muthoi*) to Callicles (526d3–4, 527a5–6).

Socrates indicates why a *muthos* is inferior to a *logos*. Calling a statement a *logos* implies that it is true, whereas calling it a *muthos* impugns its veracity.<sup>13</sup> It is, however, noteworthy that in this passage, the very *same* statement may seem *to* Socrates to be an account, and *to* Callicles to be a myth. Socrates finds truth in the traditional stories, whereas, he suspects, Callicles does not. Socrates in the *Gorgias* thus implies the following principle:

- (1) S is an account to X only if S is true for X. S is a myth to X only if S is false for X.

If we assume relativism, then whether S is an account or a myth depends on X’s beliefs. That is, S is true for X if and only if S believes that X, and S is false for X if and only if X believes not-S. However, if we assume that relativism is mistaken so X’s belief that S is true only if S is true (and X’s belief that not-S is true only if S is false), it seems reasonable to infer that:

- (2) S is an account only if S is true. S is a myth only if S is false.<sup>14</sup>

This implies that myths are totally devoid of truth and epistemic value. But within the context of the *Gorgias*, Socrates may be appealing to Callicles’ understanding of myth rather than his own.

## 5.4 Myth in the *Phaedo*

Socrates evinces a more positive attitude to myth in the *Phaedo*, beginning with the turn to poetry in the opening scene. Most of the *Phaedo* is devoted to Socrates’ thesis that “a man who has truly spent his life in philosophy is probably right to be of good cheer in the face of death and to be very hopeful that after death he will attain the greatest blessings yonder” (63e8–64a2). Progressing through a series of arguments and objections, Socrates concludes with a final argument that “the soul is most certainly deathless and indestructible and our souls will really dwell in Hades” (106e8–107e1). This argument is preceded (95e–100e) by a critique of Presocratic materialist and mechanistic explanations of the cosmos and of human thought and action. A fully satisfactory explanation would identify the cause of why things turn

out the best for each thing and for the common good of all. Socrates concedes that he cannot provide such an explanation but must instead employ a second-best explanation of laying down hypotheses, in particular, that there are Forms (for example, the Beautiful itself) and that things have the properties they do because they partake in the Forms. Socrates' two interlocutors cannot show his arguments to be unconvincing, but Simmias demurs: "in view of the importance of the subject of our arguments (*logoi*) and my disesteem for human weakness, I am compelled still to be somewhat unconvinced about what we have said." (107a9–b3) Socrates agrees, adding, "...our first hypotheses must be examined more clearly, even if they are convincing to us. And if you analyze them sufficiently, you will, I think, follow the argument (*logos*) as far as a human is capable of following it; and if this becomes clear, you will inquire no further" (107b5–9). Socrates then contends that it is just to think that if the soul is deathless, it requires our care not only for our lifetime but for all time, and that anyone who does not care for his soul faces the greatest peril. He then describes what "is said" to happen to the soul after death (cf. 107d4, 5).

The myth of the *Phaedo* is peculiar in that Socrates combines it with a fanciful geographical account, which "someone" has convinced him is true (108c8), including a description of fiery rivers flowing beneath the earth. The myth concludes with the triumph of the philosopher: "Those who have purified themselves sufficiently by philosophy live forever in the future without body; and they reach even more beautiful dwelling places which it is hard to reveal, nor is there enough time now to do so. Because of the things we have described, Simmias, one must do everything to share in virtue and wisdom in one's life, for the reward is noble and the hope is great" (114c2–9). At the conclusion of his narrative, Socrates says,

It is not fitting for a thoughtful person to insist that things are as I have described them, but I think that it is fitting and worth risking for a person who believes that it is the case—for the risk is noble—that this or something like this is the case concerning our souls and their dwelling places, since the soul is evidently immortal, and a person should repeat these things to himself as if it were an incantation, which is why I have been prolonging the myth (*muthos*). (114d1–8)

This caveat recalls, and perhaps explains, the guarded language at *Gorgias* 523a1–3.<sup>15</sup> Socrates carefully distinguishes the content of the myth from that of the preceding argument: "that these things or something like them are the case concerning our souls and their dwelling places, since the soul is evidently immortal. . . ." The dependent clause states what Socrates' argument has established: that the soul is immortal.<sup>16</sup> The myth describes the condition of the soul and its location after death. The conclusion is that "one must make every effort to share in virtue and wisdom in one's life, for the reward is beautiful and the hope is great" (114c7–8). Socrates uses very guarded language: "these things or something like them" hold of the soul afterward (114d2–3). By his own admission, Socrates cannot establish that the cosmos has a non-mechanistic, value-laden structure. The myth thus supplements the argument, asserting something which Socrates is in no position to prove.<sup>17</sup> In contrast to the *Gorgias*, Socrates here unabashedly describes his narrative as a myth, but he does not claim that it yields knowledge: "it is not fitting for a thoughtful person to insist that things are as I have described them." That is, a myth does not yield the



certainty which one demands of rational argument. Instead, Socrates says that it is “fitting and worth risking” to believe that these things or something like them is the case. Where argument leaves off, one can only take the “noble risk” of accepting the myth. By repeating the myth, like an incantation, one can work up one’s courage to take the noble risk, the leap of faith in the rewards of philosophy.<sup>18</sup>

Socrates’ characterization of myth in the *Phaedo* raises some questions. Above all we may ask whether it is fitting for a philosopher to take the noble risk described by Socrates. The “true philosopher” as described in the *Phaedo* uses “pure thought (*dianoia*) alone” as he “tries to track down each reality pure and by itself” (65e7–66a3). His reasoning (*logismos*) must be uncontaminated by sense-perception or desire, both of which depend on the body (cf. 65c2–9). But how can the true philosopher believe a myth which is not supported by pure reasoning? It might be suggested that to take the noble risk of believing the myth is an act of bravery befitting a philosopher. Socrates has claimed that only the philosopher is truly brave: “it is fear and terror that make all men brave, except the philosophers” (68d11–12). But he has also argued that true bravery, like any other genuine virtue, depends entirely upon the philosopher’s wisdom (*phronêsis*): “with this we have real bravery and moderation and justice and, in sum, true virtue, with wisdom, whether pleasures and fears and all such things be present or absent. Exchanged for one another without wisdom such virtue is only an illusory appearance of virtue. . .” (69b1–7). These remarks suggest that the noble risk envisaged at the end of the *Phaedo* would not be an exercise of genuine bravery, and that it would not be an appropriate act for the philosopher.

Granted that Socrates gives every indication that he is prepared to take the risk of believing the myth, is he not stepping outside his role of philosopher in making such a leap of faith? Could this be the point, after all, of Socrates’ recurring dream: that, in the face of death, he should practice some art other than philosophy?

## 5.5 Myth in the *Republic*

Plato seems to make some progress on this question in the *Republic*, a dialogue throughout which myths play an important role. The myth of Er which concludes the *Republic* is foreshadowed by Cephalus’ preoccupation with myths about Hades and postmortem punishment (I 330d7). Earlier in the dialogue Socrates discusses the sorts of myths which are permissible in his “mythical” ideal city (cf. VI 501e4). He is here concerned with whether traditional myths such as those attributed to Homer and Hesiod should be used as educational myths for children in the ideal city (cf. II 378e–379a, X 606e–607c).

Socrates’ argument relies on a distinction between two types of falsehood. He first remarks that musical education uses two sorts of stories (*logoi*), true and false (II 376e11). Children should learn false stories first: “Don’t you understand that we first tell myths (*muthoi*) to children? These are false, on the whole, but they are in a way also true” (377a4–7). He lays down a law that poets and speakers should be prohibited from propagating false myths about the gods (379c2–7, 380c6–9).

Socrates goes on to examine the idea that a story may combine true and false elements. He contrasts it with “true falsehood,” in which one is willing “to tell falsehoods to the most important part of himself [i.e., his soul] about the most important things. . . [i.e.,] about the things that are” (382a7–b5). True (i.e., pure) falsehood or “ignorance in the soul of someone who has been told a falsehood” is distinguished from “falsehood in words (*logoi*),” which “is a kind of imitation of this affection in the soul, an image (*eidōlon*) of it that comes into being after it and is not an entirely unmixed falsehood (*ou panu akration pseudos*)” (382b7–c1).<sup>19</sup>

Socrates does not advocate banning all lies from the ideal city. On the contrary, he notoriously argues that “falsehood, though of no use to the gods, is useful to people as a form of drug” (III 389b2–4). The lower citizens must not be permitted to lie, but the rulers in some cases may lie to the subjects for their benefit and that of the city as a whole (389b2–c6). A striking example of this is the myth (*muthos*) of the metals (415a2), which Socrates describes as a “noble falsehood” (414b9–c1): namely, the rulers were formed out of gold, their auxiliaries of silver, and the workers of bronze. But the lie resembles a child’s tale which contains a grain of truth (cf. II 377a): the citizens differ from one another in natural ability, so that it is the role of the educational system to determine who is naturally suited to which social role (III 415c1–2).<sup>20</sup>

Although the myth of the metals is an educational myth, Socrates’ defense of it may shed light on the more philosophical myths. It suggests a way to understand Socrates’ aforementioned remark in the *Gorgias* that his story about the afterlife is a myth to Callicles but an account to Socrates (523a1–3). The story resembles the myth of the metals in that it contains true elements and it engenders virtuous beliefs. The proposed solution requires that we understand “myth” and “account” each in different senses:

- (3) S is a pure myth only if S is wholly false.
- (4) S is a partial myth and a partial account, only if S is partially false and partially true.
- (5) S is a pure account only if S is wholly true.

In the *Gorgias* Socrates may be understood as imputing to Gorgias the view that his narrative is a myth in the sense of a *pure* myth, and thus wholly false and devoid of epistemic value. Socrates himself asserts that his narrative is an account in the sense of a *partial* account containing true elements. But when he calls his narrative a *muthos* in the *Phaedo*, he means that it is a *partial* myth containing false as well as true elements. His positions in the two dialogues can thus be reconciled.

Let us now see what light this may cast on the myth of Er. This follows an argument that the soul is immortal (X 608d3–611a3). The argument in brief is that a thing can only be destroyed by its own natural evil, for example, the body by disease. But the natural evil of the soul is vice which cannot destroy it. Nor can the soul be destroyed by the evil proper to the body. Socrates declares, “Let’s either refute our argument (*logos*) and show that we do not argue finely, or as long as it is unrefuted let’s never say that the soul is destroyed by a fever or any other disease nor by killing, not even if the body is cut up into tiny bits” (610a10–b3). Hence, the

soul is indestructible and immortal. Socrates states confidently that this argument (*logos*) and others compel us to believe that the soul is immortal (611b9–10).

Yet Socrates implies that the argument itself depends upon a deeper understanding of the truest nature of the soul (611b1). “It isn’t easy for anything composed of many parts to be eternal if it is not fashioned with the finest composition, as the soul now appeared to us” (611b5–7). The true nature of the soul can be grasped only through philosophical reasoning (*logismos*) (611b10–c4). We shall then see that the soul resembles the sea god Glaucus, whose primal nature is hard to discern because original parts have been broken off and others added. “Then we’d see what its true nature is and be able to determine whether it has many parts or just one and whether or in what manner it is put together” (612a3–5).

Having shown to his satisfaction that justice itself is the best thing for the soul and having solved the problem of Gyges’ ring (612a5–b5), Socrates finally argues that justice also has consequences for the just and unjust person both during human life (612b7–614a4) and after death (614a5–621d3). Hence, he says, “These things must also be heard, if both [i.e. the just and unjust persons] are to receive in full what they are owed by the argument (*hupo tou logou*)” (614a7–9).<sup>21</sup> This leads him to the myth of Er, whose soul left his body and traveled with the souls of the dead and observed them receiving judgment and choosing future lives (621b8).<sup>22</sup> The myth confirms that a proper choice of lives requires rational thought (*nous*):

[H]e will be able, by considering the nature of the soul, to reason out which life is better and which worse and to choose accordingly, calling a life worse if it leads the soul to become more unjust, better if it leads the soul to become more just, and ignoring everything else. We have seen that this is the best way to choose, whether in life or death. (618d5–e4)

The myth assures us that philosophy alone will bring us a smooth heavenly journey after death as well as happiness in mortal life (619e2–5, cf. c6–d1). Socrates concludes that the myth of Er

was saved and not lost, and it would save us, if we were persuaded by it, for we would then make a good crossing of the river of Forgetfulness, and our souls wouldn’t be polluted. But if we are persuaded by me, we’ll believe that the soul is immortal and able to endure every evil and every good, and we’ll always hold to the upward path, practicing justice with wisdom (*phronēsis*) in every way. . . . Hence, both in this life and on the thousand-year journey we’ve described, we will do well. (621b8–c6, d1–3)

The myth will be our salvation, if we are persuaded by it that a life guided by rational thought (*nous*) is always the best, both in this life and the next.

The myth of Er may be viewed as a noble falsehood, a fanciful mixture of myth and account. But it does not compel rational assent in the manner of an unrefuted argument. We might suppose that myth resembles mimetic poetry in a way, in that it appeals to a lower part of the soul (cf. 606a2–c4). Unlike a bad poem, which weakens reason’s rule over the soul, a good myth persuades the lower part to serve reason. The myth, on this interpretation, encourages the spirited part to be brave on behalf of philosophy. This is certainly, on Plato’s view, an important function of myth.<sup>23</sup>

But, if this were its sole contribution, its value for the philosopher would be purely instrumental; it would lack any epistemic validity for philosophers—even if the myth served to promote the cause of philosophy. But this would leave philosophers with the paradox that they require something beyond philosophy (that is, a myth) in order to have faith that philosophy is really all they need!

## 5.6 Myth in the *Phaedrus*

To appreciate the contribution of the myth of the *Phaedrus* to these issues, we must consider its context. Both the myth and the preceding proof of the immortality of the soul occur in a speech (244a3–257b6), which Socrates attributes to Stesichorus (a sixth-century lyric poet), which seeks to prove that love (*erôs*) is “a sort of madness given us by the gods to ensure our greatest good fortune” (247b7–c1). The speech contains a demonstration (*apodeixis*) which “will be unconvincing to the clever but convincing to the wise. Now, we must first understand the truth about the nature of the soul, divine or human, by recognizing its affections and acts” (245c1–4). This argument (145c5–246a2) has a syllogistic form<sup>24</sup>:

- (1) What is essentially self-moving is immortal. (245e2–3)
- (2) The soul is essentially self-moving. (245e3–246a1)
- (3) Therefore, the soul is immortal. (245c5, 246a1–2)

The idea that the soul is self-moving and the source of all other motion is an important advance in Plato’s psychology. For one thing, it sheds valuable light on the crucial but unexplained premise of the *Phaedo* argument that the soul is a special bearer of the Form of Life which cannot admit the Form of Death (105d3–e6, cf. 103e2–7). Unfortunately, however, Socrates provides no convincing defense of this crucial assumption. He introduces the premise in an almost diffident manner: “Since that which is moved by itself has been pronounced immortal, one is not ashamed to call this very thing the essence (*ousia*) and account (*logos*) of the soul” (245e2–4).

There is some indication that the myth in the *Phaedrus* provides support for this premise. For, immediately after arguing that the soul is self-moving and hence immortal, he says,

Enough concerning the soul’s immortality. But concerning its form (*idea*) the following must be said. To say what sort of thing (*hoion*) it is, would be altogether a long and divine narrative, but to say what it is like (*eoiken*) would be a human and shorter one. So let us speak in the latter way. Let it be likened then to the connate power of winged yoked horses and a charioteer. . . . (246a3–7)

Later Socrates refers back to this as a myth: “Just as at the beginning of this myth (*muthos*) we divided each soul into three parts, two kinds being like horses and a third form being like a charioteer, and now again let us allow these things to stand” (253c7–d1). These two passages taken together imply that the myth tells what the soul is like. The point of the myth of the *Phaedrus* is to construct a simulacrum of the soul which corresponds to it in important ways but not in every detail. The myth departs from a pure account (*logos*) which must not only distinguish the soul into

three parts but accurately represent these parts. The myth of the *Phaedrus* is a story about what the soul is not literally (a charioteer with two winged horses) intended to describe what it is like. It is thus a mixed myth, combining falsehood with truth, like the noble falsehood of the *Republic*.

Although the parallel between the chariot of the *Phaedrus* and the tripartite soul of the *Republic* is evident, they serve very different functions. The *Republic* does not speculate about the true or primal nature of the soul. The *Phaedrus* myth takes a further step by projecting, in effect, the tripartite structure of the living soul onto the soul itself, regardless of whether it is embodied. The myth offers various clues as to why these three parts must be present. In order to move itself, the soul must have its source of movement as an essential constituent: hence the winged horses. But unlike the gods, the souls are capable of both virtue and vice, so that they must have a “mixed” rather than pure source of movement: hence the good and bad horses (246a7–b4, 253d1–5). This is central to the theme of the speech, for the soul is thus capable of carnal love as well as a higher “Platonic” love. This also explains why it was possible for souls to fall from their original blessed condition resulting in the cycle of reincarnation.<sup>25</sup> The soul also requires rational thought (*nous*) as its steersman: hence the charioteer (247c7–8). The wings which are “akin to the divine” enable the chariot to ascend to heaven (246d6–e4). The soul loses its wings when it eschews the knowledge of true reality and relies on opinion instead. It falls to earth and “takes on an earthly body, which then seems to move itself, due to the power of this soul. The whole combination of soul and solid body was called ‘living thing’ [or ‘animal’], and acquired the designation ‘mortal’” (246c3–6). The embodied human soul experiences the conflicts among its three parts described in the *Republic* (cf. 253e5–257a2). Finally, the myth agrees with the other myths regarding the importance of philosophy: “Only a philosopher’s thought (*dianoia*) grows wings, since its memory always keeps it as close as possible to those realities by being close to which the gods are divine” (249c4–6). In summary, the myth contains an image which is a representation of the true nature of the soul. The myth is continuous with the everyday experiences of mortal humans and provides a way of integrating moral and religious beliefs.

The myth thus plays an important role in persuading us to accept a crucial premise of the argument. But the myth is not a *logos*: it does not state what the soul is but instead describes what it resembles. A *logos* would presumably analyze the soul into parts in such a way that we would know what the parts are, how they are combined, and why they necessarily constitute a self-moving whole. This would provide what Socrates aspires to in the *Phaedo*: an analysis of our assumptions which is sufficient to produce clarity (107b4–9).<sup>26</sup>

But the *Phaedrus* does suggest why the philosopher may need myths. Every argument about the immortality of soul must rest on assumptions, and the conclusion is absolutely incontestable only if we know the premises for certain. If we cannot give a compelling *logos* for the premises, we must accept them on some other basis. The *Phaedrus* shows how a powerful myth can persuade us to accept such a premise.

A story or picture may depict a difficult fact which we are not (yet) able to analyze logically. This does not mean that the premise can only be accepted on faith,

or that philosophy must come to a permanent impasse. It is prudent for the philosopher to be guarded about such myths, to say, “Perhaps it has a degree of truth in it, but perhaps it has led us astray” (265b6–8, cf. *Phd.* 114d–1). Yet the myths can also encourage the philosopher to continue striving for a deeper understanding of the soul.

## 5.7 Conclusion: Myth and Philosophy

Plato’s four main eschatological myths seem to suggest a problem about the relation of myth to philosophy. In each case Socrates alleges that he has heard the myth from someone else and is merely repeating and interpreting it. He denies that he is *muthologikos* or capable of fashioning myths on his own (*Phd.* 61b5). After delivering the speech attributed to Stesichorus in the *Phaedrus*, Socrates ironically claims, “I can’t remember at all because I was completely possessed (*enthousiastikon*)” (263d1–2). This may be an ironic way of establishing “distance” between Socrates and the narrative to follow (see Murray 1999, p. 256). But it may also be a way of suggesting that myths are exogenous to philosophical reasoning. However, each myth also agrees with Socrates’ belief that the soul of the philosopher will fare better after death than the soul of the non-philosopher. This is a belief which Socrates is unable to justify by means of rational argument, although he does argue that the soul continues to exist after death. This seems to present a problem: Socrates is encouraged to practice philosophy by myths which he is unable to generate himself as a philosopher. Does this suggest that the philosophical life is not self-sufficient? It is, of course, possible that Socrates in the *Phaedo* is being merely ironic in denying that he is *muthologikos*. But if this were the case, it would lead to a dilemma: is Socrates capable of making myths insofar as he is a philosopher, or has he acquired this capacity in some other way, for example through divine inspiration? If Socrates’ capacity to make myths were the result of his philosophical skill, then a myth would seem to be a mere allegory or story contrived to illustrate a truth he has arrived at independently through rational argument. Socrates sometimes recommends that educational myths be used in this way—for example, the myth of the metals in the *Republic*—in order to persuade non-philosophers to accept a true belief.<sup>27</sup> But this seems an unlikely interpretation in the case of the *Phaedo*, where Socrates gives every indication of being totally committed to pressing the argument as far as he can. If Socrates has an argument that the soul of the philosopher will fare better after death, it would be out of character for him to withhold it from his interlocutors. It is hard to see, however, what sort of argument he could advance for such a thesis based on what he has said so far in the *Phaedo*. But the other horn of the dilemma—that Socrates has acquired the capacity to make myths through non-philosophical means, for example through poetic inspiration or divine madness—leaves unresolved the original difficulty: that the philosopher as such is unable to assure himself that he is leading the best life.

Myth and reason seem especially at odds in the *Gorgias*, where Socrates refuses to call his description of postmortem punishment and reward a *muthos*. Rather,

it is an account (*logos*) for Socrates because he believes it is true, and a myth (*muthos*) for Callicles who does not believe it. The implication is that to call a description a myth is to imply that it is false. In the *Timaeus* Socrates seems to draw a similar distinction, when he says Critias' story about Atlantis is not "a fabricated myth (*plasthenta muthon*) but a true account (*alêthinon logon*)" (*Ti.* 26e4–5). Socrates' position in the *Gorgias* is paradoxical, however, in that he denies the label "muthos" to a narrative which is obviously a myth. Moreover, he claims that his narrative is a *logos* even though he apparently cannot defend it through philosophical argument.

The *Republic* takes an important step toward solving this problem by allowing that a myth may combine true with false beliefs.<sup>28</sup> It was suggested previously that Socrates uses the word "muthos" in different senses in the *Gorgias* and *Republic*: for a pure myth, which is totally false, and for a partial myth, which combines true and false beliefs. In the *Gorgias* Socrates denies that his narrative about postmortem judgment is a pure myth, but in the *Republic* he implies that the myth of Er is a partial myth. For the myth of Er contains a claim—that the soul of the just person, and especially the philosopher, will receive benefits in the afterlife—which Socrates believes to be true but which he is not able to prove through rational argument. Because the myth packages this belief with other attractive beliefs about the afterlife, the myth can be used to persuade us to believe in postmortem rewards. Because the myth involves non-rational persuasion, the implication is that it appeals to non-rational parts of the soul, for example through imagery. But this still leaves unresolved the problem of the *Phaedo*: Socrates appeals to certain beliefs which he is unable to prove by means of rational argument, so that he must have recourse to a non-philosophical means such as poetry or prophecy in order to justify his philosophical mission. The apparent implication is that philosophy and the method of rational argument is ultimately insufficient. Socrates' invocation of myth seems to suggest a mysterian view that reality is finally inaccessible to philosophical rational inquiry.

The *Phaedrus* offers a solution by suggesting that philosophy and myth can be closely allied activities: that is, the *muthos* supports the *logos* by providing a way of persuading someone to accept an unproven premise. In the *Phaedrus* the argument for the soul's immortality proceeds from the unproven premise that the soul is essentially self-moving. The function of the myth, which uses the image of two winged horses and a charioteer, is not to reveal the true nature of the soul but only "to say what it is like" (*Phaedrus* 246a5). The myth thus illustrates, and motivates acceptance of, a premise for which no proof is yet forthcoming.

The *Phaedrus* thus shows how philosophers can legitimately avail themselves of philosophical myths. If a philosopher offers a proof for a thesis—for example, that the soul is immortal—the proof must proceed from at least one premise or hypothesis. To establish the ultimate premise of such an argument—for example, concerning the true nature of the soul—would require "an altogether long and divine narrative" (246a4–5). The implication is not that the narrative is in principle beyond the reach of philosophers, but that it would require more time than they have available.<sup>29</sup>



Myth provides a way for philosophers to persuade themselves, as well as others, to accept such a premise. As a partial myth, in the sense of the *Republic*, it can contain a kernel of truth, insofar as it describes a likeness of the soul.

It does not follow that any particular premise is ultimately unknowable. For although a myth may point to a truth which is beyond the philosopher's ken at a particular time, it may be possible for the philosopher to demonstrate the proposition after further reflection. A comparison of the arguments of the *Phaedo* and the *Phaedrus* indicates the progressive character of philosophical reasoning. The following diagram indicates how a philosopher might resort to myth at a given time without conceding a mysterian view of the soul. The arrows indicate the progress of a philosopher in establishing ever more basic and more certain premises to support a conclusion which initially the philosopher only believes to be true:

Proposition	Time 1	Time 2	Time 3
Conclusion: The soul is immortal.	Proven by <i>logos</i>		
↓			
Premise: The soul is the bearer of life.	Assumed, based on <i>muthos</i>	Proven by <i>logos</i>	
↓			
Premise: The soul is essentially self-moving.		Assumed, based on <i>muthos</i>	Proven by <i>logos</i>
↓			
Premise: e.g., The essence ( <i>ousia</i> ) of soul of X			Assumed, based on <i>muthos</i>

Here the same proposition may be at one time supported by a *muthos* and later proven by means of philosophical reasoning. The rather cryptic reference in the *Phaedrus* to an “altogether long and divine narrative” (246a4–5) may correspond to Socrates' claim in the *Republic* that dialectic may ultimately enable the philosopher to finally achieve an irrefutable account (*logos*) concerning a self-evident first principle: “Dialectic is the only inquiry that travels in this way, doing away with hypotheses, to the first principle (*archên*) itself, in order to be secure (*bebaiôsêtai*)” (VII 533c7–d1, cf. 534b8–c5).<sup>30</sup>

The distinction in the *Phaedrus* between a description of what sort of thing the soul is—i.e., its essence (*ousia*) or true nature (*phusis*)—and what it is like, seems to correspond also to the distinction made by Timaeus between a secure (*bebaios*) account and a merely likely account (*eikôs logos*). In the *Timaeus* the two kinds of account differ in terms of their objects: “Accounts of what is stable and secure and transparent to understanding are themselves stable and unshifting” (*Ti.* 29b6–7). An account is called “likely” (*eikôs*) because it is an account of a likeness (*eikôn*). A likely account stands to a stable and secure account as being stands to becoming, or truth to conviction (29c2–3). Timaeus cautions, “Don't be surprised then, Socrates, if it turns out that we won't be able to produce accounts on a great many subjects—on gods or the coming to be of the universe—that are completely and perfectly



consistent and accurate” (29c4–7). Instead we must be content with the likeliest account that we can find, recalling that we are merely human. “So it is fitting for us to accept the likely myth (*eikōs muthos*) about these things and seek nothing further” (29c7–d3). The interpretation of the phrase “*eikōs muthos*” is the subject of controversy. Scholars disagree over how to understand both “*eikōs*” (does it mean “probable” or “reasonable”?) and “*muthos*” (does it mean a literal or a metaphorical story?).<sup>31</sup> These difficult issues, bearing on the purpose of Timaeus’ cosmological narrative, transcend the main topic of this essay. But the passage *Timaeus* 29b1–d3 does seem relevant to the statement of *Phaedrus* 246a3–7 that the role of a myth is to say what a thing is like (*eoiken*). Moreover, the two dialogues share an important mythical detail concerning the representation of the soul, for Timaeus says that the demiurge “mounted each soul into a carriage, as it were, and showed it the nature of the universe” (41e1–2).

The *Phaedrus*, when read in conjunction with the *Republic* and *Timaeus*, reveals a place for myth in philosophical inquiry. A philosopher could arrive at secure and certain knowledge of a thesis, e.g., that the soul is immortal, only if he could derive it from a certain starting point, e.g. knowledge of the true essence of the soul, or perhaps from knowledge of the Form of the Good. Given their limitations, mortal philosophers have to content themselves with myths they have received from poets or prophets, but which can serve as indispensable tools. By representing what the soul is like they can enable us to adopt true beliefs (or likely accounts) when the process of justification reaches a temporary stopping point.

Plato’s philosophical myth can be compared to the use of metaphors at the outer frontiers of speculation in modern physics, for example, of particle-waves, quarks, super-strings, and the like. At a certain stage we may be unable to explain the true nature of a theoretical entity, and must content ourselves with saying “what it is like” in order to proceed with an account of higher-level phenomena. Thus Plato can make philosophical use of myths and poetry even though he denies that the poet as such has any “independent authority vis-à-vis the philosopher” (Redfield 1994, pp. 44–5). Perhaps this is the point of Socrates’ paradoxical claim that he is “not *muthologikos*.”

From the fact that myth comes into play at the frontiers of philosophy, representing truths that exceed its grasp, it does not follow that any particular fact is in principle unknowable. For although myth points to a truth which is beyond the rational grasp of the philosopher at a given time, it may be possible to demonstrate this truth at a later time. Yet given the limited duration of human life,<sup>32</sup> philosophical myths will remain necessary for mortal philosophers as way stations on the search for wisdom.<sup>33</sup>

## Notes

1. Santas (1979, p. 33). In this book and an influential series of articles, Santas shows that Socrates defends his claims by means of conceptual analyses and rigorous arguments which can be reconstructed in terms of modern formal logic.

2. Edelstein (1949, p. 463) goes too far, however, when he declares, “The early dialogues, with the exception of the *Protagoras*, do not contain any mythical tale at all.” At *Apology* 40c–41b Socrates considers what would follow if the myths about Hades were true. “What would one of you give to keep company with Orpheus and Musaeus, Hesiod and Homer? I am willing to die many times if that is true.” The translations of Plato’s text are from Cooper (1997), with occasional terminological changes for the sake of consistency.
3. In the *Phaedrus* after Socrates relates “what I’ve heard the ancients said” about the Egyptian god Theuth, Phaedrus replies, “Socrates, you’re very good at making up stories (*logoi*) from Egypt or wherever else you want!” (275b3–4).
4. Friedländer (1969, p. 176) argues, “When Plato employs a myth in the later dialogues, this means he cannot express himself in any other form.” Cf. Elias (1984). In an influential early treatment, Stewart (1905) remarks that Plato’s myth “bursts in upon the Dialogue with a revelation of something new and strange.” He argues that “Platonic myth awakens and regulates Transcendental Feeling... which manifests itself as a solemn sense of Timeless Being—of ‘That which was, and is, and ever shall be,’ overshadowing us with its presence.” This Transcendental Feeling involves “imaginative representation” of the ideas of God, soul, and cosmos, which Stewart regards as Platonic precursors of Kantian “Ideas of Reason” that transcend possible sense-experience (pp. 22, 25, 44–6).
5. “We cannot tell (nor could Plato himself have told) where the figure or myth ends and the philosophical truth begins.” This remark of Benjamin Jowett serves as an epigram for Murray (1999), who argues that myth plays an integral role in Plato’s philosophy, so that the two become indistinguishable. Similarly, Rowe (1993, p. 11) cautions against “supposing that [Plato] would have drawn as sharp a distinction between his story-telling and his argumentative (‘philosophical’) mode as we may be tempted to do ourselves.” Again, Morgan (2000) speaks of “the dynamic interpenetration of myth and philosophy.” According to Morgan, “philosophical myth is rational, is deployed as a result of methodological reflection, and is a manifestation of philosophical concerns” (pp. 2, 7). See also Dalfen (2002) and Schmitt (2002).
6. I will not speculate about how the historical Socrates may have viewed myth based on the evidence of Plato, Xenophon, and other ancient sources. Nor does space permit discussion of the use of myths by Plato’s characters other than Socrates, for example in the *Protagoras*, *Statesman*, *Timaeus*, and *Laws*, or by Socrates himself in other contexts (for example, the aforementioned myth of Theuth in the *Phaedrus*), except when these shed light on Socrates’ eschatological myths. Finally, my focus here is on the relation of myth to philosophy. Annas (1982) compares the myths in the *Gorgias*, *Phaedo*, and *Republic* in terms of their content, and offers a persuasive argument that they reflect a development in Plato’s views about the afterlife. While Annas’ study provides valuable background, my main concern here is specifically with the relation of *muthos* to *logos* as expressed by Socrates in connection with myths about the afterlife.
7. Brisson (1998, pp. 141–52) provides valuable information concerning the occurrence in the Platonic corpus of “*muthos*,” derivatives of “*muthos*,” and compounds of which “*muthos*” constitutes the first term.
8. The terms are contrasted by Socrates at *Grg.* 523a1 and *Ti.* 26e4–5, as mentioned below, and also distinguished in a different way by Protagoras at *Prot.* 324d6 (not discussed in this essay).
9. Compare *Laws* IX 872d7–e2 regarding postmortem punishment: “The *muthos* or *logos*, or whatever one should call it, has been stated clearly [as handed down] from the ancient priests.” Here the Athenian Stranger is speaking, however, and his usage may reflect special considerations in legislating for the second-best constitution of the *Laws*.
10. Edmonds (2004, pp. 161–71) notes that it can be misleading to place too much weight on Plato’s usage of “*muthos*” and “*logos*.” Detienne (1986) however goes arguably too far in concluding that “*muthos*” refers to objects so innumerable and heterogeneous that Plato has no coherent theory of myth (cf. p. 91).
11. On the three categories of myth, see Morgan (2000, pp. 161–4).

12. Another translation (depending on how one construes *hōs*) is “for I shall tell you what I am about to tell you in the belief that it is true.”
13. Dodds (1959, pp. 376–7, n. on 523a2) rightly emphasizes this point, but he also interprets the *logos* as expressing in imaginative terms a “truth of religion” in contrast to a “philosophical truth.” I see no basis for this in the text, including 527a which he cites.
14. The “only if” is intended to leave open that possibility that some true statements are not accounts, and that some false statements are not myths.
15. See Dodds (1959, p. 376).
16. The clause contains *phainetai* with the participle *ousa*, indicating that the soul has been shown to be immortal rather than that it merely appears to be so.
17. I am indebted to Kevin Crotty for this suggestion.
18. In contrast, Morgan (2000, pp. 199–200) understands the risk to be “that the listener might think the suggestive nature of myth can replace dialectic.” It is difficult however to see a basis for this interpretation in the text.
19. Gill (1993) argues persuasively that Plato does not view myths as “fictional” in the modern sense, that is, as non-factual narratives which it is inappropriate to characterize as “true” or “false” in the literal sense.
20. Socrates tries to persuade the rulers, auxiliaries, and citizens “that all our training and educating of the people were things that they imagined and that happened to them as it were in a dream but that in reality at that time they were down within the earth being molded and fostered. . .” (414d1–7). In the best case the rulers themselves will be persuaded, but failing that the rest of the city (414c1–2). This noble falsehood seems to differ from another medicinal lie, the rigged lottery described at V.459c8–460a10, where the philosopher kings are supposed to lie to the auxiliaries about how marriages are arranged.
21. Morgan (2000, pp. 206–7) rightly emphasizes this reference to the *logos*.
22. Socrates distinguishes the myth of Er from the apologue of Alcinous, i.e. Books IX–XI of Homer’s *Odyssey*, which includes the journey of Odysseus to the underworld (614b2). As Halliwell (1993, p. 170) remarks, “after the critique of poetry earlier in the book, Plato is here signaling the difference between his own myth, which he offers for its philosophical symbolism, and the false myths of Homer and other poets.” Cf. Murray (1999, p. 257).
23. Edelstein (1949, p. 477) argues that “through the myth the inner core of man’s existence receives the commands of the intellect in terms that are inadequate to its irrational function.” The sole function of myth, according to Edelstein, is to “speak to man’s passions,” by enhancing courage and allaying fear.
24. For detailed analysis and reconstruction of the argument see Bett (1999) and Robinson (2001).
25. In contrast, Guthrie (1971) argues that the *Phaedrus* does not depict the soul as essentially tripartite. On his view the soul is originally pure and simple and becomes tripartite only after it has fallen. This is hard to square with the fact that the souls are likened to chariots *before* they fall, which is explained in terms of the chariot simile (248a1–b5). Guthrie’s interpretation leaves unexplained why the souls fell in the first place. Guthrie (pp. 241–2) views the “motion” of unfallen souls and gods as an unfathomable mystery for Plato. These issues are resolved more satisfactorily by the interpretation of Frutiger (1930, pp. 91–3) which Guthrie rejects: namely that the tripartite chariot of the *Phaedrus* represents a radical alternative to the monistic psychology of the *Phaedo*.
26. Annas (1982, p. 128) argues that the myth of the *Phaedrus* makes an advance over that of the *Phaedo* in that it offers a more coherent representation of reincarnation and of the ultimate salvation attained by the philosopher. For example, the disembodied state of philosophers cannot be described in terms of the geographical myth of the *Phaedo* (114c2–6). Similarly, Annas (p. 136) argues that the *Phaedrus* offers a more promising account of the ultimate destiny of philosophers than does the *Republic*.
27. Socrates dismisses allegorical explanations of traditional myths at *Phaedrus* 229c6–230a7 as distractions from his philosophical mission.

28. Cf. *Rep.* II 377a5–6 and 382b9–c1, cited above. These passages present a problem for Brisson (1998, p. 137), who contends that *muthos* for Plato is “unfalsifiable discourse,” in contrast to *logos* which is “falsifiable discourse.”
29. Cf. *Ti.* 28c3–5: “To find the maker and father of the universe is a job, and even if I succeeded, to declare him to everyone is impossible.”
30. Following the cave simile, Socrates says, “The god knows whether it’s true. But this is how things appear to me. . . .” (517b6–8). This suggests the cave simile serves as a philosophical myth pointing to the Form of the Good.
31. Cherniss (1944, pp. 421–31) and Tarán (1971) argue that the *Timaeus* creation myth is non-literal, whereas Vlastos (1965) argues that it is literal. On the meaning of “*eikôs muthos*” see Burnyeat (2005) and Rowe (2003). On the relation between the *Phaedo* and the *Timaeus* see Sedley (1989).
32. *Phaedrus* 246a4–6 seems to imply that a nonmythical explanation of the nature of the soul is in principle possible but “it would take an exposition which is altogether and in every way divine as well as long.” As Szlezák (1999, p. 97) remarks, Plato’s discerning readers “understand that what is unproven by the myth not only needs proof but is capable of being given it.” Szlezák seems, however, to underestimate the difficulty of providing such a “divine” exposition. *Phaedo* 108d4–e2 also mentions the constraint which human finitude places on philosophical inquiry. Socrates later adds that even a myth can require more time than we mortals have available (114c6).
33. Previous drafts of this paper were presented at a conference at the University of Arizona in 2002 (where Jennifer Baker was the commentator), at a conference of the Society for Ancient Greek Philosophy at Fordham University in 2003, and at a conference in honor of Gerasimos Santas at Pyrgos, Iliias, Greece. I am grateful for the criticisms at these conferences which enabled me to revise the paper extensively. I also thank Pamela Phillips for corrections on the penultimate draft.

## Bibliography

- Annas, J. 1982. “Plato’s Myths of Judgment.” *Phronesis*, 27, 119–43.
- Bett, R. 1999. “Immortality and the Nature of the Soul in the *Phaedrus*.” In *Plato 2: Ethics, Politics, Religion, and the Soul*, edited by G. Fine, 425–9. Oxford: Oxford University Press.
- Brisson, L. 1998. *Plato the Mythmaker*. Chicago: University of Chicago Press.
- Burnyeat, M. 2005. “*Eikôs Muthos*.” *Rhizai: A Journal for Ancient Philosophy and Science* 2: 143–65.
- Cherniss, H.F. 1944. *Aristotle’s Criticism of Plato and the Academy*. Baltimore, MD: Johns Hopkins University Press.
- Cooper, J.M., ed. 1997. *Plato: Complete Works*. Indianapolis: Hackett Publishing Company.
- Dalfen, J. 2002. “Platons Jenseitsmythen: Eine ‘neue Mythologie’?” In *Platon als Mythologe: Neue Interpretationen zu den Mythen in Platons Dialogen*, edited by M. Janka and C. Schäfer, 214–30. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Detienne, M. 1986. *The Creation of Mythology*. Translated by Margaret Cook. Chicago: University of Chicago Press.
- Dodds, E.R., ed. 1959. *Plato: Gorgias*. Oxford: Clarendon Press.
- Edelstein, L. 1949. “The Function of the Myth in Plato’s Philosophy.” *Journal of the History of Ideas*, 10:463–81.
- Edmonds, R.G. III. 2004. *Myth of the Underworld Journey: Plato, Aristophanes, and the “Orphic” Gold Tablets*. Cambridge: Cambridge University Press.
- Elias, J.A. 1984. *Plato’s Defense of Poetry*. Albany, NY: State University of New York Press.
- Friedländer, P. 1969. *Plato: An Introduction* (2nd ed.). Translated by Hans Meyerhoff. Princeton: Princeton University Press.
- Frutiger, P. 1930. *Les Mythes de Platon*. Paris: Librairie Félix Alcan.

- Gill, C. 1993. "Plato on Falsehood—Not Fiction." In *Lies and Fiction in the Ancient World*, edited by C. Gill and T.P. Wiseman, 38–87. Austin: University of Texas Press.
- Guthrie, W.K.C. 1971. "Plato's Views on the Nature of the Soul." In *Plato 2*, edited by G. Vlastos, 230–43. New York: Doubleday.
- Halliwel, S. 1993. *Plato: Republic 10*. Warminster, UK: Aris & Phillips.
- Morgan, K.A. 2000. *Myth and Philosophy from the Presocratics to Plato*. Cambridge: Cambridge University Press.
- Murray, P. 1999. "What is a *Muthos* for Plato?" In *From Myth to Reason: Studies in the Development of Greek Thought*, edited by R. Buxton, 251–62. Oxford: Oxford University Press.
- Partenie, C., ed. 2004. *Plato Selected Myths*. Oxford: Oxford University Press.
- Redfield, J. 1994. *Nature and Culture in the Iliad: The Tragedy of Hector*. Durham, NC: Duke University Press.
- Robinson, T. 2001. "The Argument for Immortality in Plato's *Phaedrus*." In *Essays on Plato's Psychology*, edited by E. Wagner, 323–33. Lanham, MD: Lexington.
- Rowe, C.J. 1993. *Plato: Phaedo*. Cambridge: Cambridge University Press.
- Rowe, C.J. 2003. "The Status of the "Myth" in Plato's *Timaeus*." In *Plato Physicus: Cosmologic e antropologia nel Timeo*, edited by C. Natali and S. Maso, 21–31. Amsterdam: Adolf M. Hakkert.
- Santas, G.X. 1979. *Socrates: Philosophy in Plato's Early Dialogues*. London: Routledge & Kegan Paul.
- Schmitt, A. 2002. "Mythos und Vernunft bei Platon." In *Platon als Mythologe: Neue Interpretationen zu den Mythen in Platons Dialogen*, edited by M. Janka and C. Schäfer, 290–309. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Sedley, D. 1989. "Teleology and Myth in the *Phaedo*." *Proceedings of the Boston Area Colloquium in Ancient Philosophy* 5: 359–83.
- Stewart, J.A. 1905. *Myths of Plato*. London: Macmillan.
- Szlezák, T. 1999. *Reading Plato*. London: Routledge.
- Tarán, L. 1971. "The Creation Myth in Plato's *Timaeus*." In *Essays in Ancient Greek Philosophy*, edited by J.P. Anton and G.L. Kustas, 372–407. Albany, NY: State University of New York Press.
- Vlastos, G. 1965. "Creation in the *Timaeus*: Is It a Fiction?" In *Studies in Plato's Metaphysics*, edited by R.E. Allen, 401–20. London: Routledge.

## Chapter 6

# Is the Prudential Paradox in the *Meno*?

Thomas C. Brickhouse and Nicholas D. Smith

Few, if any, essays on the philosophy of Socrates have had a greater influence on contemporary scholarship than has Gerasimos Santas' "The Socratic Paradoxes."<sup>1</sup> Prior to the publication of this paper in 1964, scholars were inclined to lump together two views that are prominent in Plato's so-called early dialogues. The first is that no one wants what is bad for oneself and so all who pursue bad things do so involuntarily. The second is that because virtue is knowledge, all who do injustice and moral evil do so involuntarily. If taken to be expressions of what is essentially the same doctrine, each implies that, for Socrates, no one wants injustice in the way no one wants, say, ill health. Not surprisingly, until the publication of "The Socratic Paradoxes," few scholars were inclined to take Socrates' moral psychology very seriously, since it seems to fly in the face of the obvious truth that some people are actually quite eager to do injustice, at least when they can do it with impunity.

The genius of Santas' essay is show how Socrates distinguishes between things that are bad in the purely *prudential* sense that they are harmful to the agent and things that are bad in the moral sense that they do harm to another person. According to Santas, Socrates *assumes* that no one wants what is prudentially bad. However, Socrates recognizes that he must *argue* for the moral claim, that all harm to others is the product of ignorance, which Socrates does, principally in the *Gorgias*. There Socrates shows that doing justice is always a prudential good and doing injustice is always a prudential evil (Santas 1964, p. 149; 1979, p. 189). By distinguishing between what he terms the "prudential paradox," that no one does bad things voluntarily, and the "moral paradox," that no one commits injustice voluntarily, Santas shows that Socrates' views about human motivation in general, and especially what he says about the relationship between knowledge and virtue, are worthy of serious consideration. Today, few scholars would take issue with the way Santas distinguishes the two paradoxes. Indeed, it is fair to say that the distinction enjoys virtually universal acceptance.

---

T.C. Brickhouse (✉)  
Lynchburg College, Lynchburg, VA, USA  
e-mail: brickhouse@lynchburg.edu

Central to Santas' treatment of the paradoxes is the assumption that Socrates holds a distinctly egoistic theory of motivation, according to which "...[no one] desires things that are bad for one, that men desire only good things, and that they do what they do for the sake of what they consider to be a benefit to themselves" (ibid., p. 162). So if Santas is right, Socrates believes there is really only one kind of desire, desire for one's own good. Once the assumption is brought out into the open, it is easy to see why Socrates would think that all harm, whether to oneself or to another, is involuntary, for harm is always the product of some sort of *mistake* the agent makes, either about what is good or about what sorts of things contribute to what is good. Many scholars refer to this sort of desire for one's own benefit as "rational" in the sense that it arises from and only from the agent's conception of what is good for him. The consensus among contemporary scholars is that on this point, too, Santas is exactly right: It is axiomatic in Socratic philosophy, these scholars claim, that rational desire is the only desire human beings possess (see, e.g., Irwin 1995, pp. 75–6; 1977, p. 78; Penner 2000, p. 164; Nehamas 1999, pp. 27–58; Vlastos, pp. 148–54).

Santas argues that our best evidence that Socrates actually holds the prudential paradox is a well-known passage in the *Meno*, 77b ff., wherein Socrates challenges Meno's third attempt to define *aretê*, virtue (Santas 1964, pp. 156–7; 1979, 185–9). Underlying the argument, as Santas construes it, is Socrates' "One Desire" brand of egoism. Although we fully agree that Socrates endorses the prudential paradox, we are not convinced that he does so anywhere in the *Meno*. In this paper, accordingly, we offer an alternative and, we believe, more plausible reading of the passage. This is important not merely for an understanding of the passage itself, but also for an adequate understanding of Socrates' theory of desire. We do not believe that Socrates is wedded to the One-Desire Theory, either in the *Meno* or elsewhere, and so by showing that in the *Meno* passage Socrates does not endorse the prudential paradox, we remove one important piece of prima facie evidence that he holds the One-Desire Theory. At the conclusion of our paper, we will comment briefly on how Socrates can hold the prudential paradox and a more complex and, we believe, more plausible, theory of motivation, one that countenances both rational and non-rational desire.

## 6.1 The *Meno* Passage

Let's begin by reviewing the *Meno* passage at issue. After having been twice criticized by Socrates for failing to explain the "nature of *aretê* as a whole" (*kata holou eipôn aretês, peri hoti estin, 77a6–7*), Meno offers the following definition: Virtue is "the desire for noble things (*epithumounta tôn kalôn*) and the power to acquire them" (77b4–5).<sup>2</sup> To test the definition, Socrates' first move is to secure Meno's agreement that the noble things the virtuous desire are good things (*ta agatha*). Socrates next focuses on the first part of the definition, "the desire for good things." He would be understandably puzzled by what Meno has said, since good things are things that make their possessor better off, and so, presumably, although not



everyone is virtuous, surely, everyone, including the virtuous, has such a desire. If Socrates is right about this, the first part of Meno's definition would fail to pick out a definitional characteristic of the virtuous. So, Socrates asks Meno to clarify what he takes to be the possible objects of desire: "Do you think that there are people who desire bad things (*tôn kakôn epithumousin*) and some who desire good things, or do not all people seem to you, my friend, to desire good things (*tôn agathôn epithumein*)?" (77b7–c2).

For his part, Meno confidently rejects the latter possibility, insisting that not everyone desires good things (77c2). Socrates is right, of course, to probe further, since it is not yet clear just what Meno thinks. No one would deny that people sometimes desire bad things. But do those who fall into this group always desire what happen to be bad things only when they mistake them for good things, or are there also people who desire bad things, knowing perfectly well that what they desire is bad (77c4–5)? At this point in the argument, it is clear that Meno is confident that there are both types of people: There are people who desire bad things, because they mistake them for good things, and there are people who desire bad things even though they recognize that they are bad (77c7).

Rather than attacking Meno's distinction head on, Socrates elects to probe a bit further to see if Meno truly grasps just what he is committed to. Meno agrees that whenever someone desires something, he desires to possess it (*genesthai autô(i)*, 77c8). But he stands fast on the consequence of the position he has staked out thus far.

Socrates now turns his attention to the second group Meno has distinguished: those who desire bad things, knowing that they are bad. Socrates focuses Meno's attention on the harm those in this group allegedly do to themselves. Of course, such people are harmed by the acquisition of bad things—that is what bad things do, by definition. But do these people really know (*gignôskontes*) they are being harmed (77e5–7)? Once again, Meno also has no reservations about asserting that such people not only think they are harmed by bad things, but they also believe (*oiontai*) that those who are harmed are miserable to the extent they are harmed (78a1–2) and that those who are miserable are unhappy (*kakadaimonas*, 78a3–4). In other words, the people in the second group Meno distinguishes *understand* the consequence of the harm bad things do. They understand that anyone is made unhappy by the possession of bad things. Moreover, the exchange certainly creates the impression that Meno regards these claims as perfectly obvious. If there is trouble brewing for Meno's insistence that there are such people, Meno, it seems, doesn't see it.

According to Santas' reading of the passage, it is at this point that Socrates has heard enough to bring the argument to a close. Socrates now asks Meno, "Is there anyone who desires (*bouletai*) to be miserable and unhappy?" (78a4–5). When Meno concedes that he doesn't think so, Socrates draws the conclusion: "There is no one who desires (*bouletai*) bad things, if indeed there is no one who desires (*bouletai*) to be such [i.e., unhappy.] For what else is being miserable but having a desire (*epithumein*) for bad things and acquiring them?" (78a6–8) For his part, Meno agrees that Socrates is probably right and that no one (is likely) to desire (*boulestai*) bad things (78b1–2).

The rest of the argument falls out easily.

S: Therefore, were you saying just now that virtue is the desire (*boulesthai*) for good things and the power to acquire them?

M: I did say that.

S: So, from what's been said everyone has this desire (*to boulesthai*) and so in this respect no one is really better than another.

M: It seems so.

S: So it is clear that if one person is better than another is, it is through the power to acquire [good things]. (78b3–8)

As Santas understands the argument, Socrates has successfully defeated the first part of Meno's proposed definition, because Meno has now agreed that the desire, that is, the *boulêsis* (the substantival form of the verb "*boulesthai*"), for good things is common to everyone, but not everyone is virtuous. Thus, the desire for good things cannot be a defining characteristic of the virtuous.

The structure of Socrates' argument, as Santas understands it, can be set out as follows: Socrates gains Meno's assent to (A) "Everyone desires (*bouletai*) good things" because it is the obverse of (B) "No one desires (*bouletai*) bad things," where this is understood to mean "No one desires bad things as the intended objects of their desire," and Meno has no choice but to agree to (B), because he can hardly disagree with the proposition (C) "No one desires to be unhappy and bad things make one unhappy." Understood in this way, the argument depends crucially on the success of Meno's retraction of (D) "There are people who knowingly desire (*epithumousin*) bad things," which is the way Meno first expressed the first conjunct of the definition, because (D) (in Santas' interpretation) is taken to be the same in meaning as (E) "There are people who knowingly desire (*bouletai*) bad things, but this is the contradictory of (B), and so Meno is driven to retract (D) as a result of his assent to (A). Moreover, the success of the argument obviously requires that the sense of "desire" employed in (D) is the same sense employed in (B), for only then do (B) and (D) contradict one another. By showing that Meno, upon reflection, really believes (B) and rejects (D), Socrates has shown that even Meno accepts the prudential paradox. No one takes what is bad for him as the intended object of any desire.

## 6.2 Only One Type of Desire?

Now *if* Socrates really is committed to the One-Desire Theory, as so many scholars maintain, doubtless he has his reasons and if anyone ever tries to argue for an alternative, Socrates would wield his *elenchos* with predictably devastating results. But notice that here we are asked by proponents of Santas' reading to assume that Meno also turns out to hold the same theory, since, in the end, according to Santas, (D) is the contradictory of (B). Yet it seems obvious at the outset of the argument, anyway, that when Socrates begins to probe his interlocutor's initial endorsement of

(D), Meno clearly does *not* hold the single-desire view. After all, Meno plainly and explicitly distinguishes the desires indicated in (D) as not functioning in the same way as the aversions indicated in (B), since he is prepared to affirm at 77c7 that some of those who desire (*epithumousin*) bad things recognize that they are bad, indicating that those in the (D) group *understand* at the time they desire what they know to be bad that their possession brings misery in its wake. So plainly whatever sort of desire Meno has in mind at 77c7 when he affirms (D) is *not* the sort that *always* aims at what is good and *never* aims at things people know will bring them misery. Now Meno may not be sufficiently sophisticated philosophically to use the language of “types of desire,” “intended objects of desire,” or “descriptions under which something is desired.” But even if he is not the brightest of Socrates’ interlocutors, it is difficult to see how he could draw the distinctions he does unless he fully believes that there are people who have bad things as the *intended* objects of at least one of their desires.

Meno doesn’t tell Socrates who these people are, but presumably he is thinking about, say, certain unhealthy people who find themselves strongly attracted to rich foods even though they know perfectly well that rich foods will do them in. For these people, their desire for rich foods seems to operate *independently* of their conception of the good. These are people who seem to have a desire for pleasure (or the avoidance of pain) as such, though they know that the pleasures to which they are attracted are bad for them. Since it seems so clear that Meno is, at least initially, thinking about people who seem to have such good-independent desires, it is worrisome to have the interpretation of the passage turn on the assumption that Meno *really* thinks there are no desires that could have bad things as their intended objects.

Further doubts about the Santas reading arise when we notice that towards the end of the argument Socrates introduces a second word for desire. From the beginning of the argument at 77b2 through Socrates’ question at 78a2 about whether people who are harmed are made miserable insofar as they are harmed, different forms of “*epithumein*” are employed. But when Socrates asks rhetorically whether anyone “desires” to be unhappy at 78a4–5, he shifts to forms of another Greek verb for “to desire,” “*boulesthai*.” Now the two words are sufficiently flexible that they *could* be used interchangeably, which, of course, is what Santas’ analysis assumes. Both Plato and Aristotle, however, sometimes employ these two Greek words in restricted, technical senses. When they do, “*boulesthai*” means “to desire what is (or what one takes to be) good.” It is a type of what we are calling “rational desire.” “*Epithumein*,” on the other hand, means “to desire what is pleasurable as such.” When Plato and Aristotle employ “*epithumein*” in this restricted sense, they are plainly countenancing the existence of what is sometimes called non-rational desire, a desire for pleasure (and the avoidance of pain) as such; it is a desire produced independently of judgments about the agent’s overall good. Moreover, the employment by Plato and Aristotle of “*epithumein*” in the restricted sense shows that they recognize the possibility of motivational conflict between a desire not to possess what one cognizes as a bad thing and a desire to possess that same thing because it appears pleasant.

Those who attribute the One-Desire Theory of motivation to Socrates will be quick to point out that from the fact that Plato and Aristotle recognize different kinds of motivation it does not follow that the character Socrates in the early dialogues does. On the contrary, some of these scholars will argue that, unless we are prepared to argue that Socrates does not really hold the prudential paradox, a position for which there seems to be compelling, independent textual authority (e.g., Plato, *Prt.* 358b6–c1; Aristotle, *NE* 1145b26–27), we have no choice but to attribute the One-Desire Theory to him in as much as the prudential paradox requires that theory.

Let us set aside for the moment the question of whether the prudential paradox relies upon the One-Desire Theory and first note the fact that the Socrates of the early dialogues *also* employs both “*boulesthai*” and “*epithumein*,” and does so with surprising frequency.<sup>3</sup> More importantly, in one passage, *Charmides* 167e1–5, Socrates explicitly distinguishes between the two in terms of their objects. No one, he suggests, would think that *epithumia* (the substantive cognate of “*epithumein*”) aims at anything but pleasure and *boulêsis* (again, the substantive cognate of “*boulesthai*”) at anything but the good.<sup>4</sup> So it can’t be said that Socrates always uses the terms interchangeably and that he is oblivious to, or even philosophically opposed to, the distinction Plato and Aristotle draw between the two.

Given the possibility that Socrates might *not* be using the two terms interchangeably, and also our earlier question of why Meno caves in so quickly about (D), when only moments before he seemed to be holding a view that required there to be different sorts of desire, let’s see what happens if we don’t assume that Socrates is relying on the One-Desire Theory and instead is employing forms of “*boulesthai*” and “*epithumein*” to refer to different desires, the two desires we saw that Socrates is aware of in the *Charmides* passage referred to above. According to this second way of reading the passage, by introducing “*boulesthai*” into the argument Socrates is pointing out that when Meno asserts (D) “There are people who knowingly desire (*epithumousin*) bad things,” Meno cannot mean that there are people who desire bad things as rational desires for misery-producing things. As (B) states, absolutely no one could think it is rational to desire to be miserable and so no one could think it is rational to want what yields misery. This has the advantage of explaining Meno’s ready acceptance of (B), for it doesn’t rely on the same sense of “desire” employed in (D). (B) asserts that no one has a rational desire for bad things and (D) asserts that there are people who have a non-rational desire for what they know to be bad.

According to our alternative way of reading the passage, Meno’s assertion (D), “There are people who knowingly desire (*epithumousin*) bad things,” is not a proposition Socrates thinks he must refute. Instead, Socrates is out to show Meno that he cannot save his definition by appealing to (D). Socrates signals that he is now appealing to a different sense of “desire” by employing a different word, “*boulesthai*,” in the formulation of (B). This is fatal to Meno because, as Socrates points out at 78b4–6, again employing a form of “*boulesthai*,” the obverse of (B), namely (A), has not been falsified by Meno’s appeal to (D). That all people have a rational desire for good things is consistent with some people having a non-rational desire for things they know to be bad. Socrates can now return to the first part of the proposed definition, only this time employing a form of “*boulesthai*”: “Thus were you just

now saying that virtue is the desire for good things and the power to acquire them (*Oukoun nun dê eleges hoti estin hê aretê boulesthai te t'agatha kai dunasthai*)?" (78b3–4). Socrates is confident that Meno must agree that this is the same desire that all persons have because the desire for good things with which virtue endows its possessor is surely a desire that has as its intended object things that benefit its possessor, not a desire that has as its intended object pleasures that may or may not happen to be good.<sup>5</sup> In other words, the only sense of "desire for good things" that Meno can attribute to the virtuous is the same desire for good things all people have, namely *boulêsis*, not *epithumia*. The difference between virtuous and non-virtuous people is that only the virtuous have the capacity to pick out unfailingly what is truly good. Others are apt to want the wrong things, at the wrong time, on in the wrong amounts, and so forth.

Looking at the argument in this way, one can agree with Santas that Socrates is out to defeat the first part of Meno's proposed definition. Socrates' strategy, however, is not to reduce the first part of the definition to absurdity by showing that, as it is stated, it conflicts with something Meno himself finds more plausible. Instead, by inquiring about what Meno has in mind in saying that there are those who knowingly desire bad things, Socrates is able to show him that no one has a *rational* desire for bad things. The principal advantage of our reading, we believe, is that it does not require that Meno holds a narrowly intellectualist view of desire, indeed, one that he seems quite explicitly to reject early in the argument. If we are right, Socrates concedes to Meno that one can have a non-rational desire for what one knows to be bad for one. Thus, *for purposes of this argument*, Socrates can concede that one can even act for the sake of what one knows to be bad for one. Our point, however, is not that Socrates rejects the prudential paradox. Nor are we arguing that in the *Meno* Socrates actually endorses a theory of motivation that recognizes more than one sort of desire. We are simply arguing that Socrates says nothing in the *Meno* that commits him to the prudential paradox.

### 6.3 Non-Rational Desires in Socratic Moral Psychology

As we stated in our introductory remarks, we fully believe that Socrates does endorse the prudential paradox. But we also think that the Socrates of the early dialogues accepts a motivational theory that recognizes more than a single sort of desire, according to which it is possible to have a rational desire for benefit in conflict with a non-rational desire for pleasure. So, we clearly deny that the prudential paradox entails the one-desire theory. Part of the explanation relies on the fact that Socrates thinks that only one sort of desire, rational desire, actually motivates action. That is, Socrates thinks that whenever we *act*, we act for the sake of what we take to be our good and whenever an agent fails to do what is in his best interest, the explanation is the agent's ignorance. To this extent we fully agree with the now received view of Socratic motivation, which owes so much to Santas. How then does non-rational desire come into play in Socratic psychology?

Let us turn to the well known passage in the *Protagoras* (352b2 ff.) in which Socrates takes on the view of “most people” regarding the possibility of acting against one’s *knowledge* of what is best. It is not in dispute in the argument that some people believe that something is bad for them at one point and yet end up pursuing it anyway. What is in dispute is how to describe what takes place. “Most people” hold that such people have their knowledge overcome by some sort of non-rational desire (352b2–c2). By the end of the argument it is clear that Socrates thinks that such people have changed their minds about what they initially thought was bad for them, for “no one who knows or believes that something else is better than what he has been doing persists when he could be doing something better” (358b6–c1). So what does Socrates think causes such people to change their minds?

The answer is what Socrates refers to as the “power of appearance,” “the power of things to appear more attractive than they really are.” This is the power, according to Socrates, that knowledge, or what he calls the “craft of measurement,” renders impotent (357a5–b4). Socrates clearly believes, then, that what explains why some people change their minds and end up doing what they previously claimed they shouldn’t is that, because they lack knowledge, some object acquires the power of appearance over them, a power that knowledge would neutralize. In cases of what most people say is “acting against one’s better judgment,” Socrates thinks an object that is initially unattractive acquires the power of appearance, and while it has that power, the agent, unless prevented, will pursue the object.

If we now ask what could have caused the object to acquire the power of appearance, proponents of the One-Desire Theory have no adequate explanation.<sup>6</sup> All they can say is that initially the agent believed the object was bad for him and so had no desire to pursue it and then later the agent believed the object was good and so formed a desire to pursue it. But if this is all that Socrates can say for his side, “most people” are hardly likely to feel impressed by Socrates’ own view about the power of knowledge, for what needs to be explained is why the object at one time appeared bad and later appeared good. “Most people” have every right to insist that the change didn’t *just happen*.

We believe that the Socratic explanation is that after initially judging the object to be bad, the agent experiences the onset of a non-rational desire for the object and that it is this desire which *causes* the agent to change his assessment. It is the experience of a non-rational desire that causes the object, to use Socrates’ terminology, to acquire the “power of appearance.” If the object appearing good is available and since we always pursue what “we know or believe to be good for us,” the agent will pursue the object, thereby satisfying the non-rational desire.<sup>7</sup>

So we believe that Santas’ central insight about Socrates’ motivation is sound. For Socrates, human agents always *pursue* what they take to be their own good and avoid what they take to be bad for them. But by resting his account of Socratic motivational psychology on his reading of *Meno* 77b ff., Santas is led to the conclusion that, for Socrates, all human desires and passions are dependent upon the agent’s beliefs or knowledge about the good. In this paper we have tried to show why we think Santas is wrong about the theory of desire assumed in the *Meno* and that when we turn to the *Protagoras* we find a more complex and more plausible view of desire

at work. In arguing against Santas' reading of the *Meno*, however, we trust that we have done nothing to subtract from the enormity of the debt Socratic studies owes to his labors.

## Notes

1. Santas (1964, pp. 147–64). The same position is advanced again in Santas' *Socrates*. See also Santas (1979, pp. 183–94).
2. In constructing his argument against Meno's definition, Socrates uses forms of two verbs that refer to conation. He begins by using "*epithumein*" and then, towards the end of the argument, switches to "*boulesthai*." As we will see, Santas and like-minded interpreters take Socrates to be using the words interchangeably, which is certainly a possibility. We believe that Socrates uses the terms to refer to different types of desires and, indeed, this distinction between desires is the key to understanding the passage. Rather than attempting to win our case "by translation," however, in our summary of the argument we will translate both terms simply by "desire," noting the Greek words employed in Plato's text in parentheses.
3. In addition to the *Meno*, see, e.g., *La.* 196d6–e1; *Grg.* 493a3–5; and *Prt.* 337c1–4.
4. We owe this point to Devereux (1995, pp. 400–1).
5. This last point is confirmed a few lines later, at 78c5–d1, where Meno asserts that what the virtuous person desires are items believed to be good, such as health, wealth, and political office, items he believes make one better off.
6. This is pointed out most effectively in Devereux (1995, pp. 390–6). See also Brickhouse and Smith (2006).
7. For an excellent, full discussion of Socrates' adherence to the Two-Desire Theory, see Devereux (1995, pp. 381–408). For a discussion of how our view, which is indebted to Devereux's, differs from Devereux's, see Brickhouse and Smith (2006).

## Bibliography

- Brickhouse, T.C., and N.D. Smith. 2002. "Incurable Souls in Socratic Psychology." *Ancient Philosophy* 22:1–16.
- Brickhouse, T.C., and N.D. Smith. 2006. "The Socratic Paradoxes." In *The Blackwell Companion to Plato*, edited by H. Benson. Oxford: Blackwell Publishing.
- Devereux, D.T. 1995. "Socrates' Kantian Conception of Virtue." *Journal of the History of Philosophy* 33:381–408.
- Irwin, T. 1977. *Plato's Moral Theory*. Oxford: Oxford University Press.
- Irwin, T. 1995. *Plato's Ethics*. Oxford and New York: Oxford University Press.
- Nehamas, A. 1999. "Socratic Intellectualism." In *Virtues of Authenticity: Essays on Plato and Socrates*, edited by A. Nehamas. Princeton: Princeton University Press.
- Penner, T. 2000. "Socrates." In *Cambridge History of Greek and Roman Political Thought*, edited by C.J. Rowe and M. Schofield. Cambridge: Cambridge University Press.
- Santas, G.X. 1964. "The Socratic Paradoxes." *The Philosophical Review* 73:147–64.
- Santas, G.X. 1979. *Socrates: Philosophy in Plato's Early Dialogues*. London and Boston: Routledge.



## Chapter 7

### Gerasimos

#### [Or Seeking Freedom from the Fregean *Under the Description Methodology*]

Terry Penner

Hot, and tired, Christopher and I are just finishing our walk from the Academy to the Lyceum. We have not been drawn into truncating our journey, it is true, by the promise of some discussion with beautiful youths in a wrestling school. But I confess that my eye, fulfilling its evolutionary function, *has*, willy-nilly, selected some dark-eyed Greek beauties along the way with whom, in another life, I might well have paused for some philosophical conversation. We reach our destination, and find we have to get down on our hands and knees in order to peer under the solid fence, getting the only view of the Lyceum possible these days. Then we stand up, dust off our pants, and start back to the hotel. On the way, we find a likely looking bar with tables on the sidewalk, and stop for a little break.

The waiter arrives while Christopher is gone, and I order a pitcher of nice cold *Muthos*. As Christopher returns, I raise my glass to him, saying “We did it! Heather and Rosemary said we wouldn’t even succeed in getting up in time, let alone doing the walk. Congratulations!” I am just about to take a long satisfying draught, when Christopher shakes his head at me, amusement in his every feature. “You don’t want to do that,” he says.

I look at him, a trifle taken aback, but with the blood rushing to my head. I see what is coming, of course. As if some further explanation were really necessary, he adds, “Are you forgetting that you are going to be facing sharp questioning this morning from those two formidable Greeks: the deep-thinking Mariana and her peerless teacher Jerry—the one and only Gerasimos Xenophon Santas? Don’t you think all that *Muthos* could well lead to your being either slow or actually befuddled in your discussion with them? And won’t the result of that, from our point of view, be that the discussion of the desire for good at *Meno* 77b–78b will be set back to the *status quo ante* of 1964, when Jerry published his great classic of Socrates scholarship on the Socratic Paradoxes? Wasn’t that the article that you like to say introduced into Socrates scholarship the name of Frege which you so revere—even while you regard him as the source of most of what has gone wrong in modern analytical

---

T. Penner (✉)  
University of Wisconsin-Madison, Madison, WI, USA  
e-mail: tpenner@facstaff.wisc.edu

philosophy? And wasn't that the article that you also like to say introduced into Socrates scholarship, for such psychological verbs as 'desire,' the logical form

verbs-the-object-*o*-under-the-description-*D*?

And aren't you the one who is always saying that this logical form forces on its users the so-called 'linguistic turn,' according to which all we can meaningfully speak of as real is the references (if any) which the *sense* or *meaning* of our words and descriptions determines, and, in addition, that the only way to characterize the direction of our inner psychological states is in terms of the direction specified by means of the description *D*. . .?"

Christopher pauses before renewing the onslaught. "And aren't you always telling me that, in this language-oriented sort of account of our intercourse with reality, the only things we can ever think about, or gain even some partial grasp on, are those things our present language can meaningfully designate? Isn't it you who is always contrasting with this Plato's account of our intercourse with reality—according to which what we think about, we think about not by grasping such things as our present language succeeds in isolating, but rather by seeing *through* our present language, as if through a glass darkly, to the real things that are there even if they *should* be different from what we think they are, and even if our present language *could not in any way* triangulate them? Aren't you always saying that this is the primary and preeminent supposition of Socratic/Platonic realism?"

He pauses; then adds, "And now you are willing to see us set back to the *status quo ante* of 1964?"

Christopher sees from the humorless scowl that I inherited jointly from my Mennonite ancestors and my Scottish Calvinist ancestors that every one of his questions has scored a direct hit. So he carries on—seeing as he does that his questions, pursuing the upper reaches of the means/end chain, have prepared the way for the *coup de grâce*.

"And what will this setting us back to 1964 prove to be a means to?" He pauses, daring me to give the obvious answer. "Yes," he concludes with triumph, "a very unhappy Terry!" Then, as if one more rhetorical question is required, he adds, "And aren't you the one who is always quoting Socrates to the effect that no one ever wants to be unhappy? Isn't exactly that point explicit in this very passage of the *Meno*?"

"You're right, of course," I say, no longer scowling, but sheepish and embarrassed—indeed quite overcome, and submissively penitent, as if I have been called on the carpet before the elders.

Even now, Christopher can't let it alone. "So which act of drinking all this *Muthos* is it that you want to do, Terry? The one that you are about to do, which will prove to be the means only to your being befuddled, and so, in the end, the means only to watching our position on the *Meno* go down in flames at the colloquium this morning, and so, in the end, the means to your unhappiness? Or do you want to do the drinking of all that *Muthos* which will refresh, and lead to no ill effects today at the colloquium, and so to no detriment to your happiness?"

Christopher is plainly enjoying this. "Oh, the second one is the one you want to do? Where you drink all that *Muthos*; get befuddled; still do a good job at the

colloquium; and live happily ever after? In your dreams! That action which you say you want exists only in some fantasy-world of yours. Your problem, Terry, is that you don't want to live in the real world!"

What can I say? Imagine: reminding *me* of how opposed I am to this whole *under the description* methodology for dealing with psychological states! Reminding *me* that what an intentional action is—its identity—includes the totality of its consequences, where at least some of those consequences are what the action proves to be means and ends to, therein, in favorable cases, mirroring the means-end sequence in the intention! Either that, or, in a *failed* or *mistaken* action, as my drinking the *Muthos* would have been, *not* mirroring that means-end sequence in the intention. It's dirty pool, isn't it, pouring over a fellow all his own deepest philosophical convictions and all his own arguments? And embarrassing him, in his folly, to boot?

"*Ariston men hudôr*, my friend," he says. I sigh and call the waiter over, pay him for the *Muthos* and ask him to take it away and bring us a pitcher of cold water instead. "Enough *Muthos*," I say to the puzzled waiter, "we're sticking with *logos* for now." Somehow, the fact that Christopher is enjoying himself so much at my expense, along with my pathetic *double entendre*, revives me. (Our discussions of the *Lysis* on the walk had seemed to raise more problems than it solved. No wonder we can't get our book finished.) But the prospect of joining the fray again with those two most dear of philosophical opponents over desire for the good—and indeed seeing Jerry and Mariana again for the first time in a year or more—all that also cheers me up again.

I turn to Christopher. "I'm really looking forward to rejoining the issue with Jerry and Mariana this morning."

"You mean, over what Socrates has in mind in the *Meno* when he says that everyone desires the good?"

"Yes," I say, "or, as I think the issue is more correctly put—over what Socrates has in mind in the *Meno* when he, in effect, puts it to us that that everyone who *does* something (or tries to *get* something: 77c8) does so desiring the good.

"You've been writing me a slew of e-mails recently about the new strategy you have been hatching to lure Jerry and Mariana over to our position. So, tell me, how are you going to present our position this morning?"

"I intend to direct myself today primarily to the main philosophical issues between us and them; and to say as little as possible about the specific features of the text that divide us. The exegetical questions get so complex so quickly."

"Yes," Christopher interjects, "our 1994 piece and Mariana's 2003 reply give eloquent testimony to that."

I ignore the implicit rebuke (since I was the one who started it all). "Indeed, just before I left on this trip to for Greece, I gave up on a hugely long attempt to restate all these exegetical issues. I gave it up because I had come to think that the only result would be to put the four of us—you, me, Jerry, and Mariana—in danger of simply passing each other by in the night of details."

"I can see how that might happen," he concedes, rubbing his face; then adds "at least in this case." Christopher hates to let any fine point of interpretation go, however small it is. Before he can back out of this concession, I move on quickly.

“On the other hand, the philosophical issues have so much to do with Jerry’s and Mariana’s resistance to our account of how the *Gorgias* and the *Meno* can be so taken as to be consistent with each other, that if we can win them over on the philosophical issues, we stand a chance of talking them out of their view that makes the two dialogues *inconsistent* with each other on this central Socratic issue of desire for the good.”

“Good,” Christopher says with conviction. “I must say, ever since our 1994 paper I have found it hard to figure out how two such brilliant thinkers as Jerry and Mariana could reconcile themselves to an interpretation that makes Socrates inconsistent between the two dialogues, once we had shown how the inconsistency can be avoided.”

“Yes, I think only philosophical doubts about the position in the *Gorgias* could have made them accept so troubling a view of Socrates.”

“So, all right, Terry, how will you begin this morning?”

Pleased at this opening for rehearsing some of the main elements of my presentation this morning, I say, “I’ll begin with two statements of the issue, one simple (using *descriptions under which*), one rather more detailed (introducing means and ends).”

“O.K., so what are these two statements?”

“In the simpler version, the issue is this: Does Socrates hold what you and I say he holds, that

(R) Agents always desire *the real good*—even if that is different from what they think it is?

Or does he hold the view I attribute to Jerry and Mariana, namely, that

(Ap) Agents always desire *the apparent good*, i.e., what they *think* the real good is—which may or may not be the real good.”

“This talk of the ‘apparent good’ being the equivalent in Aristotle to the *under the description* formulation which they prefer?”

“Yes,” I confirm. “The formulation they prefer is this:

(Ud) Agents always desire what they desire, good *or bad*, *under the description* ‘good’ or *under the description* ‘the real good’.”

“But (Ap) and (Ud) come to the same thing?”

“Yes. At least once you spell out sufficiently the job the ‘apparent’ does. For the *description under which* indicates the description which the agent attributes to the thing, not necessarily how the thing *is*—hence the *apparent* good. The description tells us how the object *appears* to the agent, where this appearance might be quite false. Hence, in the schema,

A desires *o* under the description *D*,

*D* represents the *inside* of the object of desire—how it appears to the agent—while the *o* position allows us to put in anything that is in fact *true of* the object of desire—it is the *outside* of the object of desire—even if what that is contradicts how the agent sees it.”

“All right,” he says, “I’m starting to remember a lot more clearly now some of the stuff we were saying in 1994. You’re now talking about what in 1994 we called the inside-outside view of the objects of psychological states—the inside corresponding (roughly) with Frege’s *sense*, and the outside more exactly with Frege’s *reference*.” “That’s right,” I say. “But I’ll also need to elaborate a bit this morning on some of the things you were adumbrating just now.” “Such as?”

“I’ll need to emphasize some extraordinary, and, so far as I know, hitherto unremarked philosophical implications of this Fregean *under the description* methodology, which, its proponents hold, must be employed wherever the relation between a psychological state and the world is of such a sort as to admit of the introduction of error into the psychological state.”

“So what are these implications, Terry? Out with it man!”

“The object referred to outside of the agent’s mind *can* of course, and usually does, have features not covered by the descriptions in the agent’s mind. Frege is very careful to tell us in his 1892 paper on *Sense and Reference* that comprehensive knowledge of the reference would require knowledge of the sense of every expression that refers to it, and that to such knowledge we never attain.”<sup>1</sup>

“Understood.”

“But if the whole scheme is to work, these outside objects must be such that they *can* be picked out by concepts and descriptions which are true of them and which are available in our present language and culture. No room here for thinking about, say, a not fully graspable God, or about a human good (or personal good) not currently graspable from within our language and culture.”

“Ah, yes,” Christopher says, “the linguistic turn strikes again.”

“Indeed,” I say, “the Fregean inside-outside view can never allow us to refer to anything in reality which could not be picked out by our concepts and descriptions. In particular, we can’t, on this sort of methodology, ever refer to things as they are in themselves if the way it is with those things in any way conflicts with what our limited concepts and descriptions are able to pick out.”

“Yes,” says Christopher, “I see that this is your big Socratic/Platonic move against the linguistic turn’s *under the description* methodology.”

“Indeed,” I say, really stoked up by now. “On the Fregean view, the supposed real object outside (where the psychological verb ‘desires’ is read ‘transparently’ or *de re*) can only be the reference (if any) determined by the sense or meaning of our referring expressions.”

“And what about the inside, as we see it in the *description under which*? Am I right that you’re going to say that it too is distorted by the ‘linguistic turn’?”

“You are indeed. For the *description under which* that is supposed to capture the inner attitude of the desirer (where the psychological verb ‘desires’ is read ‘opaquely,’ or ‘obliquely,’ or *de dicto*) can only be given by the sense or meaning of that description. Again only linguistic determination is allowed.”

“Let me be sure I understand you here,” Christopher says. “You wouldn’t object if the *under the description* theorists said that we can’t *know* these real things, such as the Form of the Good, or the real nature of the good, which lie beyond the capacities of our language to isolate. For you and I would both agree that Socrates and Plato would in any case say that we do not have knowledge of either of these things. That is the content of Socratic ignorance, and also of Socrates’ clear insistence in the *Republic* on his inability to identify the Form of the Good, saying that all he can do is gesture towards it by means of images (similes, allegories).”

“Dead right,” I say.

“Your complaint about the *under the description* theorists is that they don’t even admit that we might be able—even more or less adequately—to *refer to* the real nature of the good, or the real nature of cancer, if our concepts and descriptions are false of these real natures, or if our present language does not so much as allow us to formulate a true description.”

“Dead right again,” I say. “Not the sort of view you would expect to commend itself to the writer of the images of Sun, Line, and Cave, which are all about humans in fact talking about things, such as the Form of the Good for which they have no description and no conceptual grasp. When anyone tells me that this Platonic view is impossible to imagine, I just ask them about the chained prisoners in the Cave. For these chained prisoners—unintelligibly so far as their own conceptual scheme is concerned—are actually talking about those who are making the sounds they hear (echoed from the rock in front of them), namely, the statue-holders under the parapet behind them; and about the statues behind them, which are the moving things in the situation (when they *think* that all there *is* that is moving in the situation is the black shapes on the rock which are in reality mere shadows of the statues).”

“Yes, yes, Terry, enough of this particular hobby horse,” he says—not unkindly suggesting that I do no better than deal with statues of horses myself. “Back to formulating the issue that divides Jerry and Mariana from you and me.”

“Fair enough,” I say.

“So let me see if I can find where we are. You’re saying that the simpler formulation of the issue is in terms of desire being for the real good, as in (R), as opposed to desire being for things under a certain description, as in (Ap) or (Ud)?”

“I am.”

“All right, what about your more detailed formulation? Why is that needed?” “The simpler formulation doesn’t make explicit that the relevant desire is in each case the desire to *do a particular action*. By contrast, the more detailed formulation I will present makes explicit the way in which we are all concerned with those desires that bring about *actions* (as in *Gorgias* 466a–468e, and as in the reference to attempts to *get* things, referred to at *Meno* 77c8). These desires are surely the only desires relevant to ethics?”

“Agreed. So let’s hear this more detailed formulation.”

“It will be well to start with an example before us of an action—an action discussed as Socrates would discuss it.”

“An example such as you construct from Socrates’ discussion at *Gorgias* 466a–468e? The tyrant who kills his annoying prime minister, thinking he *wants* to do

this, because he thinks that this killing will prove to be the best means to his own maximum available happiness, but about which Socrates says that, since it in fact turned out disastrously for the tyrant, leading to misery ever after for him, he *didn't* want to do it: it merely seemed best to him?"

"It is," I say. "Consider now how Jerry and Mariana think Socrates should have treated this example."

"How?"

"The main thing is, they would split this account of what the tyrant wants into inside and outside, and call it:

wanting to do this disastrous killing of his prime minister *under the description* 'action of removing an annoyance from my kingdom without further bad consequences,'

where this *description under which* is simply false of that action.

"Yes, I see that."

"But if we turn to how *Socrates* sees this example, it should make two things clear to us. First, Socrates thinks that every action ever done is done by the agent as a means to this further end: the agent's own maximum available happiness; and second, he thinks that if the action does not lead to this maximum available happiness, then, whether the agent thought so or not, the action was a mistake, and he didn't want to do it."

"So in this passage of the *Gorgias*, there can be no doubt that it is our (R), and not (Ap) or (Ud), that captures Socrates' thought."

"That's right. Now, Mariana agrees about the interpretation of this passage in the *Gorgias*; and Jerry has told me he agrees as well. It is in the *Meno* where Jerry and Mariana think Socrates finally came to his senses, as they would put it, and, at least momentarily, recognized as *philosophically* compelling the view that he *needed* the idea of, say, the tyrant wanting to kill his prime minister under the description 'best available means to the tyrant's maximum available good or happiness.' Here, they are apparently saying that Socrates opts for the view he rejects in the *Gorgias*. Here, they think, Socrates is forced by the philosophical truth to abandon the position in the *Gorgias*. It's this philosophical view of Jerry and Mariana, according to which the *Gorgias* was in error to say that the tyrant didn't want to kill his prime minister, which makes it necessary for us to give a second more detailed formulation of the issue between them and us."

"And this will bring in both the action done and the means-end distinction?"

"It will. As you can see, the simpler formulation, in terms of (R), (A), and (Ud), makes no reference to means and ends."

"Right. And the same is true of Jerry's example of the bloke who reaches for what is in fact the salt shaker, though he thinks it is the pepper shaker—where Jerry speaks of the salt shaker as the *actual* object of the bloke's desire, and of the pepper shaker as the *intended* object of his desire. And this way of speaking he treats as interchangeable with talk of *wanting the salt shaker under the description 'the pepper shaker'*."<sup>2</sup>



“Indeed,” I say. “Whichever way he puts it, there is no reference to the *end* to which getting the apparent pepper shaker would be a means—to pep up the food, say, which is, further, a means to making the food more enjoyable, which is, in turn, the means to living the happiest life then open to the agent.”

“But will Jerry and Mariana accept that this means-end stuff is always involved with Socrates, and in particular in *Meno* 77b–78b?”

“I can’t see that they would hesitate over the point that means and ends are, *in general*, involved in every action according to Socrates. As for this passage in the *Meno*, I think we showed pretty conclusively in our 1994 piece that it too involves the idea that in any desire to do something or to try to get something, the action desired will be a means, and the end involved will be the agent’s own maximum available happiness. Indeed, I think we may have been the first to make abundantly clear the centrality of the means-end hierarchy rising all the way to the agent’s own happiness in the discussion of desire at *Meno* 77b–78b.”

“Oh,” Christopher says, “You’re thinking, at least in the first instance, of our account of the rejected possibility that some people desire what is bad knowing that it is bad and harms. This alleged possibility is spelled out as desiring what is bad knowing that it will lead to harm and so to misery and unhappiness ever after (77d5–78a8); and *Meno* agrees it is to be ruled out on the grounds that no one wants to be unhappy ever after.”

“Yes,” I say, “that suggests clearly enough that what is in question in general is desiring to get something which is a means to benefit, which is in turn a means to happiness.”

“But we can *also* provide compelling argument about the relevance of the means-end hierarchy to what Jerry and Mariana think is the crucial five-line argument of the passage: 77d7–e4—the discussion of the alleged possibility that people sometimes desire what is bad thinking it good, at 77d7–e4.”

“Yes,” I say, “Jerry and Mariana treat these five lines as giving a self-contained deductive argument.”

“I know you have this animus against deduction, Terry (Penner 2007); but I’m not quite sure why you bring that point up just here.”

“My point is this: They treat the argument as one that can be made without the need for any further assumptions outside of these five lines—e.g., assumptions about any means-end hierarchy.”

“Yes, I see now.”

“Look, here (roughly) is the way they translate the five lines—in a translation we contest in our 1994 piece:

(J1) [People who desire bad things thinking them good] don’t desire these bad things, since they don’t know they are bad; rather

(J2) they desire things they think good, *though they are bad*<sup>3</sup>;

So,

(J3) they desire good things.”

“Yes,” says Christopher, “Here Jerry regards good things in (J3) as the *intended* object and bad things in (J2) as the *actual* object; or put in his preferred way,

Jerry treats this passages as telling us that the agent desires bad things under the description ‘good things’.”

“Right,” I say. I then show Christopher three brief passages from Mariana’s 2003 piece—with italics added by me to the six phrases containing the word “only”—and say “So, treating this as a self-contained, purely deductive argument Mariana too has it that

...[In this passage],... what Socrates is denying is that bad things are the *intended* objects of these people’s desires. For his *only basis in the passage* for saying that the people in the hypothesis do not desire bad things is the statement that these people do not know that the things which, according to Meno, they desire, are bad [J1]; and the *only relevant statement that follows from this* is that bad things are not the intended objects of their desire.

This quote continues immediately:

Further, *the only basis that Socrates has in the passage* for saying that the people in the hypothesis really desire what is good [J3] is the statement that they thought that the things, which according to Meno they desire, were good things [J2]; and the *only relevant thing that can follow from this* is that good things are *intended* objects of their desires [J3]. . . .

This now enables Mariana to defend Jerry’s reading as follows:

But Santas’ attribution of the descriptions view to Socrates in the *Meno* is based on the fact that the ignorance of the people who think bad things are good *seems to explain why they don’t desire bad things*. He makes a very persuasive point in arguing that ignorance about bad things *can only be relevant to desiring them* if it is *because* one doesn’t know them that one doesn’t desire them, which *can only be true* if we add the description ‘bad things,’ thus leaving open the possibility that one *does* desire them, under the description ‘good things’.”

Christopher looks at these passages, and himself underlines the six phrases with the word “only,” and says, “I see what you’re getting at with these six phrases. The point is that both Jerry and Mariana hold that the argument *excludes* premises from anywhere outside these five lines.”

“Exactly right,” I reply. “In particular, as I just said, no means-end assumptions. But, as you were suggesting just now, we showed in 1994 that there are means-end assumptions in the immediately preceding lines of the passage, namely, that

To desire something = to desire that one gets it (77c8–d1)

and

To desire that one gets it = to desire that one gets it thinking it a means to benefit (77d1–2 with d2–7).

And implicit here must be—as a passage later in the *Meno* (87e–89a) makes clear—that

To desire something  
 = to desire to get it  
 = to desire to use it  
 = to desire the benefit to which (wise) use of it is the means  
 = to desire happiness.”

“I see,” he says, “And this shows that it is not enough simply to speak of desiring the salt shaker under the description ‘the pepper shaker’ or of desiring bad things under the description ‘good things.’ We need to think in terms of the entire means-end structure running from the action all the way up to the agent’s own happiness.”

“Exactly right,” I say. “This should tell them that it cannot be a matter of the tyrant wanting to kill his prime minister under just any old description. It has to be a matter of his wanting to do the killing under the description ‘really best means to my own maximum available real happiness’.”

“I see that.”

“But notice also that the reference to the end of happiness and indeed to the entire means-end hierarchy appears only in the *description*. There is no reference yet to happiness as it is in reality, or to real happiness as would be best found in the tyrant’s situation.”

“All right, I see that too. The desire in question, on their view, can only be the desire to do the action which constitutes what the agent *thinks* constitutes the best means to what the agent *thinks* his own happiness consists in.”

“Good,” I say.

“So, to sum up, Terry, we were unquestionably right in suggesting that the *Meno* passage, just as much as the *Gorgias* passage, was intended to show that, for Socrates, wanting to do something was always to be handled in this means-end way.”<sup>4</sup>

“I think we were.”

“So now I think I can finally ask to hear your more detailed formulation of the philosophical differences we have with Jerry and Mariana.”

“The question that divides us from them is this,” I say, writing down my formulations on the paper table cloth. “Does Socrates hold, as we say he does, that

(Rm/e) the desire that explains the action done is the desire to do that action which is the really best means to the agent’s maximum available real happiness?

Or does he hold, as Jerry and Mariana must hold, that

(Udm/e) Agents always want to do whatever action they end up doing under the description ‘the best available means to his or her greatest available happiness’?”

“Good,” says Christopher. “I think that puts the issue the way we want. And we see from this way of putting the issue that it is not a matter of a straight fight between these two claims:

(R1) What the tyrant wanted was to do the really best action, *period*: that is, *not* killing his prime minister;

and

(A1) What the tyrant wanted was to do the action he actually did, *period*: namely, killing the prime minister.

If it were a matter of such a straight fight, it would augur badly for Socrates' position in the *Gorgias* and for our position."

"Good point," I say. "For as Tony put it to me, nearly twenty years ago now, about my account of this tyrant passage, this account *would* face the following really serious problem: How am I to answer the question:

(Q) If the tyrant didn't want to do it, why *did* he do it?"

"Yes," Christopher says, "I see that we have there a serious question. For one of your big claims was that the Socratic theory of motivation is a species of belief/desire theory. Every action, in other words, must be explained by the agent putting together a belief with a desire, with the result that the action is generated. So surely the result of such a theory, if it is to explain the tyrant's killing the prime minister, must be that it will generate, in any particular case (even where a disastrous mistake is involved) a desire to do *the actual action that was done*? And that makes it look as if it must be a desire to do the action of killing the prime minister that is involved. But then that appears to contradict the claim in the *Gorgias* that the tyrant *didn't* want to do that disastrous action."

"That's right. If we were allowed to say only either 'He wanted to do this killing, *period*' or 'He wanted *not* to do this killing, *period*,' we'd have to go with desire for the merely *apparent* good: to do this killing, *period*."

"And I suppose what you are now going to insist on is that once we bring in ends which the action is a means to, then the balance will shift towards us."

"Yes."

"Still, it has to be said, Terry, our position can still look a little shaky." "Why do you say that?"

"Because it still seems as if we are explaining the wrong action—the action which is the really best means to the really best end—namely, *not* killing—instead of the action that *needs* explaining, the killing which is merely the apparently best means to the apparently best end."

"Yes, I'm in total agreement about how the situation *looks*. Shall I tell you why Jerry and Mariana *seem* to have the advantage here over us?" "Do, please!"

"It's because, implicitly, they are modeling the process from the desire for happiness, via belief, to action, in terms of a simple syllogism:

(J1a) The tyrant, because what he wants, ultimately, is his own happiness, wants to do whatever action he will come to *think* is the really best means to what he *thinks* is the really best end, i.e., what he *thinks* constitutes his own happiness.

(J1b) He thinks this killing is the really best means to the really best end.

So,

(J1c) he wants to do this killing.

So,

(J1d) he does this killing.”

“And thinking this syllogism to be merely an alternative way of expressing the *under the description* theory, they tend to suppose that they thus have in hand the *only* way to explain the actual killing done; and that is why they conclude that it is going to have to be conceded that the *Meno* is inconsistent with the *Gorgias*.”

“That’s right,” I say. “And it seems very simple—a simple case of discovering a particular action which fits what the agent *thinks* would be best, and substituting to get, as the conclusion to this bit of ‘practical reasoning,’ that the agent *does* the action. But actually, it is a little more complex than this.”

“How so?”

“Let’s look at the process more historically (or more archaeologically), rather than in the usual logical manner, as designed to get a syllogism with the actual action done as, so to speak, its valid conclusion.”

“How do you mean, looking at it more historically rather than logically or syllogistically?”

“I’m suggesting here the following historical or archaeological model of the process *from* the generalized *whatever*-desire *to* the action—as Jerry and Mariana must conceive it:

(J2a) the tyrant, because he wants to be happy, wants to do whatever action he *thinks* fits the description ‘really best means to my real happiness.’

(J2b) The tyrant *thinks* that *this* killing is the action which fits this description.

So,

(J2c) The tyrant wants to do this killing under the description ‘really best means to my real happiness’.”

“And surely it’s very clear here,” Christopher interjects, “that Jerry and Mariana will have to admit that

(J2c) The tyrant wants to do this killing under the description ‘really best means to my real happiness’

is far more likely to fit Socrates’ conception of the process from desire for one’s own happiness, through belief, to the action actually done, than the bald

(J1c) He wants to do this killing,

which lacks any reference whatever to the *description under which*.”

“Exactly right,” I reply. “And this allows us to point out that here, there will be, even for them, a kind of ‘incoherence’ in any desire to do this killing *tout court*.”

“I see what you’re saying. Because the description under which the tyrant would be doing it would actually be false of the action he does, there will be a kind of incoherence in the actual desire.”

“Yes,” I say. “That feature—the incoherence—does not show up in the simpler, bald version (‘He wants to do this killing’) which they might have thought they could represent themselves as operating with. But they are in no position now to jettison the *under the description* methodology. So now we can ask the following question:

How do they get from the desire to do this action under a certain description in (J2c) to the doing of the action?”

“At this point,” I say, “what Jerry and Mariana will have to do, on this model, will be to borrow some terminology from Łukasiewicz, and say something like this: that

(J2d) Since the moment for action has now come, the tyrant, rightly or wrongly, thinks it acceptable to *detach* the action *tout court* from the *description under which*, and simply do the action *tout court*.

Hence,

(J2e) The tyrant does the action *tout court*—and regardless of what descriptions might be true of it.

“All right,” Christopher says, “I see that this gives a plausible historical model of how the action is generated according to the *under the description* theory. It involves this hitherto unnoticed—and somewhat dubious—detachment of the bald

He wanted to do the killing  
from the more nuanced

He wanted to do the killing under the description ‘best means to my happiness’.”

“Precisely,” I say, “and do you also see what I think is the most striking feature of this mode of explanation?”

“What’s that?”

“Let me proceed in a backward order. Does the tyrant here do the wrong action? An action that does not fit the description under which he wants to do it?”

“Yes, he does. For he fails to notice that the action he does is not in fact the best means his description requires.”

“So we see an error in the choice of means apparent at stages (J2c) and (J2b)?”

“We do.”

“And will Jerry and Mariana have to grant, as we grant, that there is this error in the tyrant’s choice of means?”

“I’d certainly have thought they would have to. For, as you pointed out just now, not only the end of the tyrant’s own happiness, but the entire means-end hierarchy must, for Jerry and Mariana, show up only in the *description under which*—the description which, *ex hypothesi*, the tyrant attributes to the action.”

“But, now, wouldn’t the error in choice of means be sufficient to explain the mistaken action, without a further error in choice of end?”

“I suppose it would.”

“But can Jerry and Mariana, on their theory, stop there?” “How do you mean?”

“Don’t they have to push the error back into the initial desire to do whatever action is the really best means to the tyrant’s own real happiness?”

“Ah! I *think* I see what you’re saying.” “So tell me.”

“You’re pointing out that they don’t treat the generalized *whatever*-desire in (J2a) as we do, as a desire for a certain means to the tyrant’s own real happiness. Rather they drive what we think is merely an error in the choice of means *believed to be best*, all the way back into the choice of end *desired*, so that, not only is the *belief* misguided, so is the generalized *whatever*-desire and the *desire for the end*.”

“Hmmm. . ., that seems a bit fast, Terry.” “How so?”

“I mean, couldn’t they say that the tyrant is still right about the nature of happiness which he wants for himself, and wrong merely about the means?”

“Certainly, that could happen, Christopher. But not, I think, in the general case. Consider this very example of the tyrant. Evidently, if this tyrant thinks that killing his prime minister will be to his benefit, he will have a conception of the nature of happiness according to which happiness can be gained by harming other people.

Not so, Christopher?”

“Yes.”

“But then what the tyrant *attributes* to happiness, and what his description of happiness picks out *cannot* be the real nature of happiness, according to Socrates.”

“Agreed.”

“Here, since the happiness is also merely under a description, they are not talking about the tyrant’s real happiness, but about what the tyrant *thinks* his real happiness is—which will in this case evidently be quite another thing.”

“Yes, I see that now,” he says. “For them, it’s not just a matter of his desire for his real happiness being frustrated by his ‘screwing up’ in the wrong choice of means. It’s a matter also of the very desire for happiness being ‘screwed up’ by a further wrong choice of end—of what constitutes his happiness—with the result that in these kinds of cases, the tyrant doesn’t desire his own real happiness, but merely his *apparent* happiness.”

“Right,” I say. “Whereas the Socrates of the *Gorgias* has it that people always desire their real happiness.”

“Got it,” he says. “So what should our strategy be now?” “What do you think?” is my reply.

“Shouldn’t we be pointing out that we can describe this process from generalized *whatever*-desire to action no less adequately? And that we can do that without any need to drive the error back into the *whatever*-desire? And without any need to drive the error into the very desire for the tyrant’s own happiness—or to decide that the tyrant *doesn’t* desire his own real happiness?”

“Clearly,” I say. “Absolutely.”

“So deploy for me the corresponding historical model for *our* theory.” “It will go like this:



(T1) The tyrant wants to do whatever action is the *really* best means to his own *real* happiness, even if that differs from the action he may *think* is the really best means.

So,

(T2) The tyrant, whether he knows it or not, wants to do the particular action which *is* the really best means to his real happiness [in this case, the *not* killing of his prime minister].

Unfortunately,

(T3) The tyrant, because of various false beliefs he has, *misidentifies* the action which is the really best means [the *not* killing, which in this case, unbeknownst to him, is what he actually wants to do] with this particular action of killing his prime minister.

So, then, as a result of this misidentification, the tyrant's desire in (T2) gets 'screwed up' with the following (incoherent) result:

(T4) As the tyrant sees it, he wants to do that particular action which is the really best means to his own real happiness [in fact, some particular *not* killing], i.e., this action of killing his prime minister.

Then, the moment of action being upon him, and given that he has misidentified the action which constitutes the really best means to his happiness (and which is the action he *wants* to do) with this disastrous act of killing, we now have an explanation of the fact that

(T5) The tyrant does this particular action of killing the prime minister.

"I see," Christopher says triumphantly. "The point, once more, is that the desire that explains the killing is not the desire to kill the prime minister *tout court!* What (T4) tells us is that

(T4\*) the desire that brings about the killing is the desire to do the action [not killing] which is the really best means to his real happiness, once, due to his false beliefs, the tyrant misidentifies this really best action he wants to do with this killing of the prime minister."

"Yes," I say. "Excellent! So to sum up, the desire that brings about the killing is the desire to do an action which is quite other than the killing, but which, because of the tyrant's false beliefs, then comes to be *diverted*, or *misdirected*, to this action of killing. It is the *beliefs* leading to the misidentification that are mistaken. There is no mistake in the end the action is designed to secure. That, as I see it, is our position."

“What you’re saying,” Christopher says, “is that in (T4\*), we have an account that suffices to explain the action done, but which also preserves the account in the *Gorgias* according to which it merely *seemed* best to kill the prime minister. The killing was not the action the tyrant *wanted* to do: merely the misbegotten result of a diverting of the tyrant’s actual desire to do the particular action which would make him happy to the disastrous killing.”

“That’s right,” I say, “And here’s a way to see that we are right to say this. We simply ask: Which is the explaining desire?”

(Dr) The desire to do the action which constitutes the really best means, even if that should be different from this (disastrous) act of killing?

Or

(Da) the desire to do this killing, even if it is disastrous?”

“You’re suggesting that, whatever the explaining desire is,” Christopher says, “it is not just

(Dw) the desire to do this killing whether it turns out well or badly.”

“That’s right.”

“And, as I see it, this disarms Tony’s question. For the question:

(Q) If the tyrant didn’t want to do this killing, why did he do it?

gains a foothold only because we are thinking that, as in (Dw), whether the killing turns out well or badly is irrelevant to the desire that brought it about. Tony’s question only seems compelling if we disregard the history of the action from

initial desire for whatever action constitutes the best means,  
through  
identification of the action to be done,  
to  
the action itself.

Both models must locate the error somewhere in that history of the production of the mistaken action.”

“Yes, I think that’s right,” I say. “What we have done is to show that such errors are always compatible with its being universally true that, in every action, the agent’s end is his or her own real happiness. That is why *Meno* 78a6 tells us, in effect, that good, bad, wise, and foolish people all desire the same thing in every action. They do not differ from each other in their fundamental desire for their end. Only in their beliefs.”

“And this is your explanation of the claim that Virtue is Knowledge—no wicked, perverted, or self-indulgent ends, as in Aristotle: just mistaken means.”

“And that is also what I think lies behind Socrates’ claim that the unexamined life is not worth living.”

“Yes clearly. But notice one more thing, here, Christopher.” “What’s that?”

“It concerns this misidentification of the action the agent wants.” “So what about it?”

“This misidentification is exactly parallel to the non-Fregean misidentification of Theaetetus with Theodorus when Theodorus is seen in the distance.”

“Look, Terry: I know you are broaching something very important to you here. Indeed, you’re forever pouring this stuff about the Wax Tablet into my ears. But I am not sure I am up to your now starting to discuss the question why Plato, in the *Theaetetus*, rejected what you call the Fregean Wax Tablet Model for false belief.”

“Just laying down a marker here, Christopher.”

“Fine!” he says, with just a trace of irritation at me, for throwing out *way* too many such philosophical markers in my conversation—a fault I acknowledge, and which others have described as my having *too many bees in my bonnet*. “Let’s get back to the present passage,” he continues. “I see now how this historical model of yours explains *Meno* 77d7–e4. It should be read, not as Jerry and Mariana wish to read it, as a self-contained deductive argument, but in terms of other assumptions in the argument—means-end assumptions, going right back to 77c8.”

“Agreed.”

“The result is that the argument taken as a whole, from 77c8 through d4, then reads, with our 1994 punctuation at 77e2, as follows.

(TC1) To want something = to want to get it, and use it wisely, that use being the best means to the agent’s own benefit, [and so to his or her real happiness].  
So,

(TC2) People who don’t know that what they desire is bad [and leads to harm— which they *don’t* desire] don’t desire what is bad;

(Why do people not desire harm? Because, as we see in 77e5–78b2, that leads to unhappiness.) Rather (at best)

(TC3) they desire what they *think good*.

But (that won’t work either, since)

(TC4) what in this case they think good *is* bad [and, once more, leads to harm, which they *don’t* desire: so they don’t desire what they think good either].

So, it remains that

(TC5) they desire what is [in fact] good [the real good, as at 77b8–c1 with c2–3, which deals with what is *in fact* bad, though it may be *thought* good].”

“Exactly right,” I say.

“So,” Christopher continues, “since our way of explaining the tyrant’s mistaken action of killing is *possible*, we now see how we can maintain that the tyrant does not want to do this action of disastrously killing the prime minister that he does. And since, on our reading, the *Meno* is consistent both with the [supposedly] earlier *Gorgias* and with the later *Republic* 577d (where the tyrant also *does nothing that he wants to do*); while the account of Jerry and Mariana is not; surely it’s conclusive that our reading is the way to go.”

At this point, I check myself from saying that I am not so sure myself about Jerry and Mariana regarding our argument as conclusive. Instead, I decide simply to say, “Well, let’s see how it works on Jerry and Mariana this morning.”

“Speaking of this morning, . . .” Christopher says, looking at his watch. . . . His voice tails off. Then, “Oh Sugar!” he says, with his usual quiet vehemence.

“What’s the matter? Is it time to go?”

“Not only is it time to go,” he says. “There isn’t even time to go back to the hotel, shower, and change, let alone to get some breakfast. We’re going to have to go as we are, and simply hope Heather and Rosemary will have nicked some fruit and Athenian pastries for us from the breakfast buffet.”

“Yes,” I say, “as all husbands know in their hearts, that was the real back-story of Adam’s lame rationalization in Eden.”

I leave a tip for the waiter, wave *efcharistó* to him, and we head off towards the National Gardens.

As we walk along I talk about what to say if Jerry and Mariana say that it is impossible to desire *except* under a description—that it is impossible *ever* to talk about desiring—from the inside—the real good. (The persistent line for those who have taken the linguistic turn.) I try to get Christopher to listen to my suggestions—which I foolishly pile onto him without any effort at a half-way intelligible explanation—that

- if we can use indexicals (*I, he, she, here, now, there, this, her*, and so forth) and proper names to refer to, and have in mind, the things that are actually there outside of our minds, even if they differ from the descriptions we would give of them; and
- if we can *have in mind* the real nature of cutting even if that differs from what we *think* it is, or even if it differs from what our conventions (or meanings) determine it to be; and
- if we can *have in mind* the famous ‘man in the corner drinking a martini,’ even if he is in fact a secret teetotaler, and is drinking water from a martini glass; and
- if we can *have in mind* and *pursue*, as what we desire, what is really best, even though that differs from what we think it is, and even while we are always, as the *Republic* has it, in perplexity about it, and unable to grasp sufficiently what in the world it is;

then, obviously, we can urge that

it *is* possible to desire what is *really* best for those we love, and not merely what we would ourselves at the time *describe* as ‘really best’ for them, or even what *those we love* would *describe* as really best for them;

and, similarly, we can urge that

it *is* possible to desire what is really best for ourselves, even though that differs from the *description* we would give of it.”

But Christopher ignores this ample supply of new markers, being in any case now intent on getting to the colloquium on time. (Nor will he listen, I know, if I try to put to him my point that the whole idea of deliberation is that it shall be about the real things that are there—and not simply about our descriptions may happen to pick out (if anything). For only in this way can we attempt to realize our deliberations in actions which deal with *those same real things that are there*.)

So it is that we forge ahead at the double. That’s O.K. for Christopher: he is in twice as good shape as I am. Fortunately, it isn’t many minutes before we see the Zappeion straight ahead and to the right. We see Heather standing at the bottom of the steps, haloed in the sunlight, and frowning at us, as if we were two naughty boys. (Not far off, that.) “What have you two been doing?” she mouths at us—plainly mortified, since, fierce guardian of Christopher’s interests that she is, she always *gets him to the lecture on time*. As we get closer, we can make out more of the mouthed remarks intended for us only: “Where did you walk to? The Peiraeus?” We cringe, and start up the path towards the steps. People are beginning to appear from the doors. First down the stairs is George’s brother, the affable Athanasios. “Oh, I’m glad you got here,” he says to us. “We were worried that something had happened to you.”

“Well, in a way it did,” says Christopher.

“Is Mariana recovered from the jet lag yet?” I ask.

“Oh, didn’t you know?” uncle Athan replies, “Mariana didn’t come in the end.”

“Oh dear, I hope nothing is wrong?”

“No,” he says, “George made an announcement at the beginning of this last session. It appears that an opportunity came up for her back in Del Mar which she just couldn’t turn down. There’s this young athlete who has just burst onto the scene who wants her to train him for Olympic dolphin riding; so she decided she had enough on her plate back in Del Mar.” I see that Christopher is as relieved as I am that it is nothing more serious, even if sorry not to have her there. “Jerry’s here though,” Athanasios continues, “and on good form as always. So now that we know you two are here, George tells me we should have plenty to discuss in your session.”

People begin to emerge from the doors in greater numbers. We can see Jerry now, in conversation with a group of people. He positively explodes with loud laughter at the humor he obviously sees in questions Mike and Alex are peppering him with. Further over, coming from another entrance, we can see Yuji in animated conversation with Naomi and Tony. At the top of the steps, further still to our left, we see our friend George—*autos ho Georgios*, the leader of our little philosophical band here

in Hellas—and beside him Myrtali—twin pillars, they, of the Hellenes in Southern California. Christopher and I prepare ourselves for George who is clearly going to tweak us for our missing the earlier session. With a twinkle in his eye, he will surely mumble some dry, but kindly humor at our expense. Myrtali whispers something to him, and they both look back for their friend. Jerry, meanwhile, has seen us, and breaks away from the group surrounding him and moves towards the steps. “Terry! Christopher!” he cries in his powerful, mellow *basso cantante*, startling all the bystanders. “I thought you were going to give me an easy time of it by not showing up!” and he laughs again. Meantime Rosemary and Heather each hold out paper bags reproachfully, urging us at least to have some stuff they scored from the buffet. They do not hector. But their looks tell us that they think we *really could have* made enough of the occasion to get back in time for a shower, a change of clothes, and some breakfast. I wolf down a *spanakopita*, and turn to see Jerry right behind us. I give him a bear-hug. “*Oh Sachs, mein Freund!*” “Yes,” he says, “and here are all the other *Meistersinger* with their *Lehrbuben* processing out from the lecture hall.” And some more clarion Santas-laughter bathes the assembled company. (Jerry enjoys his own wit almost as much as that of others. He does right.) We then pass some pleasant time talking, as others join us too. Someone (Fred was it? Or Hugh?) emerges from the lecture hall and starts to wave people in. Christopher and I turn to see what has happened to Heather and Rosemary. They are talking with Ann and Myrtali. We each grab one of Jerry’s shoulders, wave to our Andromaches, and escort our Achilles up the steps into the field of battle, trembling for what is to come.<sup>5</sup>

## Notes

1. See Frege (1892/1997, p. 53). In an inferior manuscript, Terry also argues that such knowledge of the sense of every expression in the language that refers to a given thing would *not* be sufficient for comprehensive knowledge of the reference, since there will always be features of the reference that the language is inadequate to express.
2. In the inferior manuscript, Terry denies this interchangeability, on the grounds that, in this example, we seem invited to suppose that somewhere else on the table there is also the pepper shaker which is what the desirer wants; but Jerry’s account is explicitly designed to allow *descriptions under which* that pick out, so to speak, things that don’t exist—an idea which Jerry finds entirely unproblematic (Santas, 1979, p. 316f, n. 22). Terry denies that there *are* any “things that don’t exist.”
3. The punctuation and translation of (J2) is contested in Penner and Rowe 1994. For their reading of the passage which Jerry and Mariana treat in terms of (J2)–(J3), see (TC3)–(TC5) below.
4. In another manuscript of this dialogue, there is a marginal note that reads as follows: “This explains why the argument considering the possibility of desiring bad things thinking them good does not occur where it ought to by Jerry’s lights, at 77c3–5; but rather only later at 77d4–e5, where it is in fact part of the discussion of the possibility of *desiring what is bad knowing it to be bad, but thinking it will benefit*. The explanation for this delay is that the lines intervening between these two passages, 77c5–d3, introduce the means-end distinction into the account of desire, so that now we have a true exposure of the hierarchical structure involved in the supposed case of desiring something thinking it good—which is all that remains once one sees that there can be no such thing as desiring what is bad knowing it is bad, but thinking that it benefits.”

5. I would like to thank my friend Christopher Rowe for his willingness to allow me to represent him as my straight man here. Though I have learned much from him, the words and thoughts attributed to him here are entirely mine, and should not be attributed to him. I would also like to thank Mariana for gracious and helpful comments (not to mention objections that tied me in a knot for a decade and more). And last, but not least, I would like to thank Jerry for more than a decade of true friendship—but also for a single (characteristically generous) comment on what I regarded as an over long earlier draft. His comment helped me see, without actually saying it, that the huge amount of textual exegesis in that earlier draft was a mere distraction. (The master teacher at work again: our graduate students should all be so lucky.) And that brought me to concentrate on the philosophical issues between us. Such issues aside, it has been an enormous comfort to me that I have always found so much agreement between Jerry and myself on matters of the interpretation of Socrates and Plato. That fact, along with our philosophical agreements and disagreements, has made my conversations with Jerry among the most enjoyable and profitable of my entire scholarly and philosophical life. What is more, I learn from him every time I reread one of his pieces. In this, I am quite sure my reaction is not unique.

## Bibliography

- Anagnostopoulos, M. 2003. "Desire for the Good in the *Meno*." In *Desire, Identity and Existence: Essays in Honour of T.M. Penner*, edited by N. Reshotko, 171–91. Kelowna, Canada: Academic Printing & Publishing.
- Frege, G. 1892/1997. "On *Sinn* and *Bedeutung*." In *The Frege Reader*, edited and translated by M. Beaney, 151–71. Oxford: Blackwell Publishers.
- Penner, T. 1991. "Desire and Power in Socrates: the Argument of *Gorgias* 466a–468e that Orators and Tyrants have no Power in the City." *Apeiron* 24(3):147–202.
- Penner, T. and C. Rowe. 1994. "The Desire for Good: Is the *Meno* Consistent with the *Gorgias*?" *Phronesis* 39(1):1–25.
- Penner, T. 1996. "Knowledge vs. True Belief in the Socratic Psychology of Action." *Apeiron* 29:199–230.
- Penner, T. 1997a. "Two Notes on the *Crito*: the Impotence of the Many, and 'Persuade or Obey'." *Classical Quarterly* 47(1):133–46.
- Penner, T. 1997b. "Socrates on the Strength of Knowledge: *Protagoras* 351B–357E." *Archiv für Geschichte der Philosophie* 79(Heft 2):117–49.
- Penner, T. 2000. "Socrates." In *The Cambridge History of Ancient Political Thought*, edited by C.J. Rowe and M. Schofield, 164–89. Cambridge: Cambridge University Press.
- Penner, T. 2002. "The Historical Socrates and Plato's Early Dialogues: Some Philosophical Questions." In *New Perspectives on Plato, Modern and Ancient*, edited by J. Annas and C.J. Rowe, 189–212. Cambridge: Harvard University Press.
- Penner, T. 2003. "The Forms, the Form of the Good, and the Desire for Good in Plato's *Republic*." *The Modern Schoolman* 80:191–233.
- Penner, T. 2004. "Towards a prudential reading of the *Crito*." In *Socrates: 2400 Years Since His Death. International Symposium Proceedings*, edited by V. Karasmanis, 13–21: July 2001, Delphi: European Cultural Center of Delphi.
- Penner, T. 2005a. "Platonic Justice and what we mean by 'justice'." [www.nd.edu/~plato/plato5issue/Penner.pdf](http://www.nd.edu/~plato/plato5issue/Penner.pdf) (76 pp.).
- Penner, T. 2005b. "Socratic Ethics, Ultra-Realism, Determinism, and Ethical Truth." In *Virtue, Norms, & Objectivity: Issues in Ancient and Modern Ethics*, edited by C. Gill, 157–87. Oxford: Oxford University Press.
- Penner, T. 2005c. "La Forme du Bien et le Bien de l'Homme: Quelques Problèmes d'Interpretation du passage 504A–509C de la *République*." In *Études sur la République de Platon, Vol 2: De la Science, du Bien et des Mythes* (a collection of papers on the *Republic* for a year-long seminar



- on the *Republic* at the Sorbonne, edited by M. Dixsaut with the collaboration of F. Teisserenc, written in French, pp. 177–207).
- Penner, T. 2006a. “The Forms in the *Republic*.” In *The Blackwell Guide to Plato’s Republic*, edited by G. Santas, 234–62. Oxford: Blackwell Publishers.
- Penner, T. 2006b. “Plato’s Ethics: Early and Middle Dialogues.” In *The Blackwell Guide to Plato*, edited by M.L. Gill and P. Pellegrin, 151–69. Oxford: Blackwell Publishers.
- Penner, T. 2006c. “Forms and the Sciences in Plato.” In *A Companion to Plato*, edited by H. Benson, 165–83. Oxford: Blackwell Publishers.
- Penner, T. 2007. “The Death of the so-called ‘Socratic Elenchus’.” The C.J. De Vogel Lecture at the 7th triennial meetings of the International Plato Society, in Würzburg, Germany, Summer 2005. In *Gorgias-Menon*, edited by M. Erler, M. Brisson, and L. Brisson, 3–19.
- Penner, T. and C.J. Rowe. 2005. *Plato’s Lysis*. Cambridge: Cambridge University Press.
- Santas, G. 1964/1979. “The Socratic Paradoxes.” *Philosophical Review* 73. Reprinted in Santas, G., *Socrates*. London: Routledge and Kegan Paul.

# Chapter 8

## Beyond *De Re*: Toward a Dominance Theory of Desire Attribution

Naomi Reshotko

### 8.1 Introduction

Well-worn examples have led many philosophers to claim that verbs expressing psychological states or activities force their objects to refer in a non-standard way.<sup>1</sup> When Oedipus married Jocasta, he married his mother. Prior to his marriage, Oedipus would have confessed to wanting to marry Jocasta, but he would have denied any desire to marry his mother. The view that psychological verbs place their objects into an intentional context (where we only need worry about how the object of the verb is conceptualized by the subject of the verb) allows Oedipus' claims concerning these desires to be taken at face value. When he kisses Jocasta, he kisses an actual physical human being who is one and the same person as his mother—he kisses his mother. This cannot be denied. But, when he *desires* Jocasta, he desires her as he conceptualizes her, or under a certain description. It has been asserted that desires take as their objects, not actual physical things, but only the conceptions that we have of those things (*apparent objects*, which might not correspond to—and therefore, not lead to—actual objects). For this reason, conclusions that would follow from premises that contain non-psychological verbs do not follow from premises that have the same form but contain psychological verbs. If it is true that Oedipus kissed Jocasta, and that she is his mother, the conclusion that he has kissed his mother is entailed. However, if it is true that he *wants* to kiss Jocasta, and it is true that she is his mother, the answer to the question of whether or not he *wants* to kiss his mother could indeed be negative.

This view can be expressed through the introduction of a terminological distinction. It is said that the object of a psychological verb must be understood *de dicto* and not *de re*. When the context dictated by a verb is not psychological, we take the object of the verb to refer to a particular state, no matter how the state is conceptualized or described. Thus, if it is stated that Oedipus has the chicken pox, it is not the case that we need to know how Oedipus conceptualizes his illness, or even whether

---

N. Reshotko (✉)  
University of Denver, Denver, CO, USA  
e-mail: Naomi.Reshotko@du.edu

or not he believes that he has it, in order to determine whether or not the statement is true. In this case the object of the verb is understood *de re*. The object refers to a specific and objective state regardless of how it is conceptualized or described. When the object of a verb is read *de re*, co-referring expressions can be substituted for that object *salva veritate*. When a verb is psychological, it dictates a context that requires that its object be read *de dicto*—as a description under which the subject understands it. In this case, co-referring terms cannot be substituted *salve veritate*. A person can want to meet the first U.S. Ambassador to France without wanting to encounter the inventor of bifocals, even though these two expressions—“the first U.S. ambassador to France” and the “the inventor of bifocals”—are co-referential.

Or so it has been argued. Herein, I challenge this view<sup>2</sup> by demonstrating that when we observe a person’s behavior in the absence of any information about what that person believes we find ourselves making justifiable hypotheses concerning what a subject desires even without knowing how she conceptualizes that which she pursues. Furthermore, in many of these cases, learning more about how the subject conceptualizes the object that she pursues will not persuade us to revise our hypothesis concerning what she wants. That is, in experimental contexts, where we assume that we know the actual parameters under which a subject is operating, it makes sense to understand all desires in a *de re* fashion. Furthermore, this indicates that there are no circumstances under which a subject is the final authority when it comes to stating what she desires—that is, no desire should be understood in a *de dicto* fashion.

Two supposed counterexamples might be raised to the claim that desires are best understood in a *de re* fashion. Counterexamples that, if legitimate, yield potential criticisms of the proposal that analysis of all desires should take place along *de re* lines. My purpose in discussing, and defending against, these counterexamples is twofold. First, the fact that such counterexamples are proposed illuminates the assumption inherent in the traditional view (the view that desire attributions must be made in a *de dicto* fashion) that subjects are the best authorities when it comes to the content of their desires—better authorities than are relevant experts, like psychologists. As I have already indicated, this assumption should be challenged. Second, I wish to concede that, while they are more successful than *de dicto* attributions, even *de re* attributions are not satisfactory for giving explanations of human behavior. To the extent that there are difficulties with explaining behavior by means of *de re* desire attributions, these difficulties indicate that we must move toward a theory that acknowledges that desires have both intentional and extensional aspects. I will propose that, in order to accurately fix the object of desire, we must move toward what has been described by Terry Penner as the *Dominance* theory of desire.

## 8.2 The Argument for *De Re*

The argument that the object of a desire is better understood in a *de re* fashion is mounted upon two foundational assumptions that I take to be uncontroversial. The first is offered by Donald Davidson and his followers<sup>3</sup>: desires (together with

beliefs) provide causal explanations for purposive behavior. The second is to be associated with behaviorism, but also functions in our everyday and common sense approaches to desire attribution: a major criterion for determining that a desire has been satisfied is the termination of some specific behavior that the desire motivated. In well-defined, observable circumstances, the fact that behavior terminates is evidence that the desire has been eliminated, which is, in turn, evidence that the situation most recently obtained was the one that satisfied the desire in question and so, was what the desire in question was *for*.

Qualifications must be made with respect to what exactly counts as extinction<sup>4</sup> and how reliable a piece of evidence extinction is. Extinction will ultimately turn out to be a blunt instrument that does not distinguish between cases in which a subject's desire is satisfied and cases in which a subject is too ignorant to realize that his desire has not been satisfied. For our present purposes, however, the fact that people typically are willing to consider extinction as a helpful indicator of desire satisfaction will suffice to show that the general assumptions of the common sense approach to desire satisfaction (1) weigh heavily against the *de dicto* theorist's supposition that subjects have an edge over observers when it comes to knowing what they desire, and (2) weigh heavily in favor of the assumption that desires have a hierarchical structure.

In fact, the assumption that the extinction of a desire-motivated behavior is the key to the determination of the object of the motivating desire is actually the assumption that desires have a hierarchical structure. We generally attribute desires to non-human animals or to pre-linguistic humans by assuming that they have some sort of overarching goal as a context for their purposeful behavior. We say that the infant who stopped crying when fed was *hungry* because we assume that the hunger caused discomfort, which in turn led to the crying. We generally assume that the infant did not want the food *per se* but wanted the comfort brought by the food. For this reason, the infant's failure to stop crying even when fed leads us to surmise that, since food did not extinguish her crying, it must not have brought her sufficient comfort—she did not want food.<sup>5</sup> This hierarchical structure is assumed any time we try to attribute a specific desire to a subject through observation alone. The essence of my conclusion will be that knowledge of a subject's conceptions is not necessarily helpful when it comes to the formation of a hypothesis concerning what that subject desires. When it comes to the formation of this sort of hypothesis, a subject can be just as misled by a lack of knowledge concerning what is an actual means to his or her ultimate goal as an observer can be. Parents are certainly accustomed to overriding their children's own—very honest—professions concerning what they want on the basis of observations concerning behavior and its extinction. It can seem transparent to an observer that a child who is convinced that she is not hungry will not be able to concentrate on anything until she eats something, or that a child who seems desperate to hang out with a certain group of peers would be thoroughly relieved if family obligations prevented him from doing so.

To illustrate this, let us look at an experimental example where we observe the lane-changing behavior of a driver. If we see the driver change to the right lane and relax and stay there, we are likely to hypothesize that he wanted to drive in

the right lane. However, if we see him change to the right lane and continue to seem frustrated, check his mirrors, and light his turn signals (exhibit “lane-changing behavior”—whatever that is) as he drives in the right lane, we are likely to hypothesize that he did not really want to drive in the right lane. In this case, our likely hypothesis would be that he wanted some further thing that he thought he could get by driving in the right lane. Taking some other aspects of his behavior (time of day, route driven) and other beliefs (that are not beliefs about what he desires) into account, we might hypothesize that what he really wanted was to get home as quickly as possible. Thus, in order to give an adequate explanation of his behavior, we would at least need to refer to whatever it is that he thought he could accomplish by means of driving in the right lane. We should regard this as the case even if it is not how the driver represents his desire to himself. This is because the hypothesis that he “wanted to drive in the right lane” does not seem to explain the matter at hand as thoroughly as a description that cites whatever further goal he was trying to accomplish at least for the simple reason that it explains less. It explains his behavior only up until he switched into the right lane. The explanation that cites his more ultimate goal in acting would explain his behavior both before and after he switched into the right lane.

Following this analysis let us explain the behavior of Oedipus without recourse to any preconceived notion concerning what Oedipus wanted. Oedipus exhibited a certain sort of agitated behavior prior to his marriage to Jocasta. Had he exhibited contentment afterward, we might rationally hypothesize that he wished to marry her. However, over the course of his marriage, his behavior is observed to grow even more agitated (he goes from mooning around before marriage to gouging out his eyes afterward). Thus, it seems that a more comprehensive explanation of his behavior would be one that can account for what he did *both* before *and* after he got married and that seems to point to some further goal that he was attempting to accomplish by means of marriage. Rather than hypothesizing that he wished to marry Jocasta (which I am happy to grant is how he represented his desire to himself) we should hypothesize that he wished for life-long happiness and, as a means toward that goal, he wished to marry someone who was, at least, not his own mother. Should someone object that scope of explanation is not the criterion that we should use in order to say which is more adequate, it will be interesting to note that this is exactly the criterion that compels objectors to develop the first apparent counterexample discussed below.<sup>6</sup>

When we are forced to consider how we make hypotheses about desires under experimental conditions, we see that, if our results are to explain behavior, they do so best once we recognize that all desires must fit into a hierarchical means-ends structure. Thus, many desires are for the means to one ultimate goal, are subordinated to the desire for that ultimate goal, and must be understood in relation to the achievement of that ultimate goal. While only a few might hold that there is only one ultimate goal—something like happiness,<sup>7</sup> even for those who hold that there are multiple, parallel, hierarchies of desire, it is still the case that most desires must be understood as desires for the means to a more ultimate goal, and these desires are subordinate to the desire for that more ultimate goal.<sup>8</sup> I claim that a

*de re* attribution is to be preferred because it is how things *actually are* with the object that is obtained (the *res*) through a given piece of behavior that determines whether or not that object, and also the behavior that led to it, are indeed a means to that more ultimate goal.<sup>9</sup> Since desire aims at what is actually a means to our ultimate goal (and not at what we merely think is a means to that goal), substitutions among terms intended to refer to those actual means are not only permissible, but necessary, within psychological contexts. For only then can the resulting desire attributions provide adequate explanations for behavior. In fact, the rules for substitution can be the same for psychological contexts as they are outside of such contexts—a means by any name is equally a means.

Of course, we must understand that when we consider practical and experimental contexts we artificially isolate one piece of the hierarchy and treat what is actually a proximate goal as an ultimate one.<sup>10</sup> In cases where behavior is not extinguished we are forced to use the more ultimate goal to give a more comprehensive explanation for the continuing behavior. Thus, it seems that, in these cases, we will have to look at the *res* that we are taking to be the more proximate object of the motivating desire in order to figure out whether or not it really is the object of that more proximate desire. The more proximate desire must be understood to be a desire for a means to a particular end.

Counterexamples have been offered to the claim that substitution allowed outside of a psychological context must be equally permissible within psychological contexts, if desires are to participate in adequate explanations of behavior (Huston 2000, p. 154). Though the counterexamples that have been offered are illegitimate and so do not affect my original analysis, an examination of these counterexamples is instructive for clarifying what needs to be put into place once *de dicto* desire attributions have been abandoned.

### 8.3 The First Counterexample

It is telling that the notion that we need recourse to *de dicto* characterizations of the object of desire in an intentional context is fueled by examples in which contemporary theorists *assume* that they know both what the subject wanted and how the subject behaved. They assume that they *know* that Oedipus both wanted to marry Jocasta and didn't want to marry his mother. Then, they ask themselves how, given this assumption, we are to explain the fact that he married someone who *was* his mother. Thus, part and parcel of this theory of intentional contexts is the thesis that agents know what they desire when they act and are reliable informants when they articulate their wishes. But, a major objective of my above argument that desire attributions should be analyzed in a *de re* fashion is to ask whether it makes sense to assume that each agent is an authority on what she desires (this was also a major objective of Reshotko 1996).

Thus, I place the problem on a different footing than that assumed by those who embrace the *de dicto* reading of desire attributions. I focus on how we come to conclusions as a result of observing behavior when we have no preconception of

what object that purposeful behavior pursues. It is for this reason that I discuss how we make hypotheses concerning the object of a desire in an experimental context. In this case, I claim we come to conclusions about what a behavior pursues on the basis of what extinguishes that behavior. I also claim that in order to do this we need to presuppose that desires have a hierarchical structure where most of what is desired is desired as a means toward some further, overarching goal. In addition, I point out that the results that we glean through this method are often inconsistent with agents' own claims about what they desire. On this basis I encourage reconsideration of the thesis that explanatory desire attributions should be formulated on the basis that agents know what they desire. The claim that explanatory desire attributions are to be understood in a *de dicto* fashion—and the counterexamples designed to show that (below)—presuppose that agents know what they desire, since they assume that it is the sense of the object of desire (the way that the agent represents the object of desire to herself) that is the true object of an intentional verb.

The two apparent counterexamples that I consider beg the question that my argument is designed to answer. I argue that we should look at how we make hypotheses concerning the object of desire when we must do so based on strictly behavioral and contextual evidence in order to hone our instincts concerning how we should arrive at these hypotheses in cases where we have a pre-experimental bias concerning what the object of the desire that is motivating the observed behavior might be. I do this in order to show that, if we do so, we will often correct ourselves in cases where we did have a preconception concerning the object that a given behavior is supposed to pursue. I am also, thereby, challenging the claim that the object of a desire should be understood as referring to the *dictum* rather than the *res*. This is because I am looking at cases where I am not taking desire claims made by the subject to be evidence concerning what the subject actually desires. To claim that we should read desire attributions *de dicto* is to claim that subjects *know* what object is motivating their own purposeful behavior. A legitimate and effective counterexample to my claims cannot be one that starts from the assumption that we know what, for example, Oedipus, desires without looking to his behavior.

Now let's look at our first proposed counterexample. In this example everything is the same as it was in my Oedipus example (above) up to the point where Oedipus marries Jocasta. However, Oedipus never finds out that Jocasta is his mother and, also, exhibits contentment after his marriage rather than an increase in agitated behavior. An objector might be misled into believing that this is a counterexample because, if we assume that we can attribute the same desire to Oedipus that I hypothesized for my *earlier* example (that Oedipus desires to marry someone who is, at least, not his mother) we can no longer explain the behavior that Oedipus exhibited after he married Jocasta.<sup>11</sup> In other words, this counterexample assumes that we all know that Oedipus didn't want to marry his mother and, since we all already know Oedipus didn't want to marry his mother, we must resort to a *de dicto* reading of what Oedipus thought he was doing when he married Jocasta if we are to explain both his agitated prenuptial behavior and his content post-marital status. But, I am not at all sure on what basis anyone observing Oedipus' behavior *in this case* can claim that Oedipus did not want to marry his mother. What I hypothesized



for *my* example is not relevant here. Our objector is observing and attempting to explain a whole different set of behaviors. Unless he can show that the desire that he attributes to Oedipus is derived from Oedipus' behavior both before and after his marriage to Jocasta, he is assuming that which he is attempting to prove. Namely, he is claiming that, in order to know what Oedipus wants, we should look at what Oedipus thinks he wants.

In contrast to my earlier example, this example is one in which we have no right to make any claim whatsoever about Oedipus' desire to marry his mother. Until further evidence indicates otherwise, there is no reason to contemplate a more ultimate goal or to move to a *de dicto* reading. It seems Oedipus wanted to marry Jocasta and he is doing just fine with the *res* that he got. However, this is where we must appreciate the limitations of the behaviorist-type assumption that extinction is an indicator of desire satisfaction. We cannot claim to know what Oedipus wanted simply because agitated behavior was extinguished and did not return over the course of his marriage; Behaviorism leaves us in a verificationist position that does not allow us to distinguish between a case where an agent has obtained what he wanted and a case where his desire motivated behavior is extinguished because of false beliefs that he maintains about the object that he obtained as a means to satisfying his desire. Still, the ease with which we adopt extinction as a criterion for assigning an object of desire—allowing it to actually override an agent's own professions concerning what she wants—shows that the common sense assumptions we apply in figuring out what someone wants—particularly in experimental circumstances—do not cohere with those made by the *de dicto* theorist and do include the conviction that desires have hierarchical structures.

## 8.4 Beyond the *De Re/De Dicto* Distinction

No one can plausibly resist my assertion that while we sometimes take a person to desire a particular state of affairs with limited knowledge of its properties, when that state of affairs obtains; it comes to be with all of its properties. However, an adherent to the *de dicto* method of desire attribution might maintain that this is irrelevant. Such a theorist will claim that it is only "*how* the agent represents the situation that explains the agent's behavior and this is gotten at through opaque readings of the agent's desires" (Huston 2000, p. 189).

In order to proceed to an even more adequate analysis of desire attribution, I grant that some aspects of agents' psychological states are important to explanations of their behavior. I do not grant, however, that agents' conceptions of their object of desire are either the *only* or the *most important* factor that must be taken into account. We cannot explain behavior using only *de dicto* desire attributions. However, the assumption that we explain behavior through either *de dicto* or *de re* analyses is mistaken. This is the failing of my preceding analysis (and Reshotko 1996). While my earlier analysis does demonstrate that *de dicto* attributions will not adequately explain behavior, *de re* readings will also fail to accomplish this. What is needed is a theory that recognizes that, certain intentions of agents, which are

inherent in how they represent every desire to themselves, force us to take account of how things in the world actually are (the *res*) when we try to understand what motivates purposive behavior.

The *de dicto* theorist and I have taken opposite sides in an argument where how things are in the world (what Penner has called the “outside” of the object) is regarded as separated from how the object appears to us (Penner 2005, pp. 161 n. 8, 186–7, calls this the “inside” of the object). We have each been mistaken in arguing that one looks to only one of these “sides” of the object of desire in order to adequately explain behavior.<sup>12</sup> What is needed is an account that takes note of the fact that agents always *intend* (inside) for how things are in the world (outside) to trump any discrepancy between how they represent an object of desire to themselves (inside) and the way the world really is (outside). Terry Penner and Christopher Rowe (1994) offer such an account. They explain as follows:

Keith Donellan has argued that when Jones uses the words “the man in the corner drinking a martini” to refer to someone who is in fact—and unbeknownst to Jones—drinking water from a martini glass, these words can, in one perfectly standard use, be taken to refer to the man in the corner drinking water from the martini glass. [To go one step further one could argue that] in such cases Jones’ *state of mind* is one of intending to refer to the man in the corner drinking water from a martini glass. We are not just speaking of a ‘transparent’ reading of “Jones intends to refer to *x*” [which would give the outside of the object]. We are speaking of the *inside* of Jones’s psychological state. It is Jones’s intention to refer to the man *as he actually is* and even if how Jones would describe the man is other than how he actually is. Jones wants *how it is with the man in question* to over-ride any errors in Jones’ conception of him. (pp. 5–6)

Penner and Rowe go on to add,

(In general, people are well aware that their conceptions, and descriptions, of people they are referring to are inadequate.) “But he’s only drinking water,” we say to Jones. “Whatever!” he replies, “You fix it up. (And when you’ve fixed it up, *that’s* the person I intend to refer to.)” (p. 6)

Penner and Rowe point out that it is a feature of Jones’ psychological state that he *intends* (from the inside) that how it actually is with the man in question (outside) dominate his manner of referring to that man (inside). Jones intends for his reference to overcome any discrepancy between how he represents things to himself and how they are. Penner and Rowe argue that this same intentional (inside) feature must be brought into play with regard to how people represent their desires to themselves.

This insight carries implications for how we should understand a desire for something that is desired as a means to some further end. We want (inside) the way the object actually is *particularly insofar as it may or may not be an actual means to the further end for the sake of which we desire it* (outside) to dominate mistakes that we have made in representing (inside) that object to ourselves as a means to that further end.

In the driver example, we hypothesize that the driver wants to drive in the lane that will get him home fastest. We hypothesize that it is on this basis that he chooses to drive in the right lane. But his desire to drive in the right lane is premised on his belief that it is the fastest lane. That is, he wants (inside) however things actually

are with respect to the speed of that lane (outside) to dominate his representation of the right lane (inside) as the lane in which he wants to drive. He wants to drive in the fastest lane *whichever one it is*, but he has mistaken the right lane for the one in which he actually wishes to drive. Oedipus wants however things actually are with Jocasta to dominate his representation of her to himself as the woman he wants to marry.

It is correct to say that in order to give a satisfactory explanation of behavior we need to look at an agent's psychological state but—and very importantly—we also need to look at how things actually are in the world. The agent's psychological state requires that we assess what he desires by looking at any discrepancy between the object he represents to himself and the way the world is (“You fix it up!” as Penner and Rowe say). The *de dicto* theorist overlooks the importance of understanding desires as descriptions of means-ends hierarchies to our explanations of behavior. This is what results in his assertion that how an agent represents the object of desire to himself is the only thing to which we need to look in order to explain that agent's behavior.

## 8.5 A Second Apparent Counterexample

A second supposed counterexample to my original argument allows us to examine further interesting issues. Here an objector might assume that he knows what the object of desire is because an experimenter has “planted” it. Six year olds are told that the Easter bunny is in the next room and it is assumed that their subsequent behavior—no matter what it is—is to be explained by citing their desires with respect to meeting the Easter bunny. Now it seems that, in this case, there is no question that the explanation will invoke the *dictum* as there is no *res* to which to refer.

However, here again, the assumption that one can know *in advance* that any behavior exhibited by the children is motivated by a desire whose object involves the Easter bunny begs the question with respect to what we should look at in order to determine what desire the children have. Until we observe the children's behavior both before and after they see what is in the next room we have no basis for making a claim. A child who skips eagerly into the next room might continue to skip and be happy even upon finding it empty—this might encourage the experimenter to revise any desire hypothesis she made based on the child's behavior before the child saw the empty room.

In this case, however, in addition to pointing out that this is not a counterexample, it is also important for me to address another point that this counterexample is designed to make. Isn't it the case that desires for non-existent objects *not only can, but also must*, be read in a *de dicto* fashion? If this is the case, what justification is there for using one theory of desire for existent objects and another for non-existent objects?

I can defend my original thesis against both of these points by saying that, where a non-existent object is wanted as a means to some further end, we still must make

*de re* desire attributions—a non-existent object cannot be the means to some further goal. The children either desire or don't desire to meet the Easter bunny on the basis of how they think that meeting the Easter bunny will fit into their overall plan for satisfaction of curiosity, fun, stimulation and happiness. That's what determines how each child reacts to the news that the Easter bunny is in the next room. Any debate between me and my objector would be limited to the issue of whether or not non-existent objects ever serve as the *ultimate* goal in a means-ends hierarchy. If they do, then according to my former principles, I should, perhaps, agree that the ultimate goal in a means-ends hierarchy must be treated *de dicto*.<sup>13</sup>

One solution to this problem is a Socratic one: Socrates thinks that all motivational hierarchies work toward happiness: all desire is for the good. And the ultimate human good is happiness. Diotima impresses this upon Socrates at *Smp.* 204e2–205a3. Our desire for our own happiness plays itself out as an effort to maximize the amount of happiness we experience over the course of a complete life.<sup>14</sup> Thus we might understand all motivational hierarchies to work toward a particular and determinate terminus, which is the maximum amount of happiness that the agent is able to experience during his life given the way the world is. This is a *res*. It is an objective goal which it is possible for us to hit or to miss. It is not an apparent object, whose identity is exhausted by that which is strictly internal and psychological. It is the real and independent, ultimate object of our desires. It exists just as do other goals to which we refer in discussing our future plans. My goal of becoming a grandmother and my goal of becoming certified in CPR are goals like this one. They are objective destinations that I can approach, at which I can arrive, and from which I can withdraw, because they are there to get closer to and farther from. The question of whether or not I have reached or grown closer to them has an objective answer that is not determined by how they or anything else appears to me. They are not physical objects that exist at the present moment in space and time. However, unlike the Easter bunny (I presume) they are something rather than nothing at all.<sup>15</sup>

## 8.6 Going Beyond *De Re*

While my original theory answers all of the problems that have been presented with respect to non-existent objects, the Dominance theory still provides the best analysis for these cases. In the case of supposed desires for non-existent objects we see that the Dominance theory handles these in exactly the same way as it handles any other desires. There is no question that subjects intend for the non-existence of objects that they take themselves to need as a means to some further end to trump their representation of those objects as existing when they take themselves to desire them. Where it seems that a non-existent object is the ultimate goal for which other desired objects are a means, the question of how we should describe the object of desire can be answered as follows: either the agent will still want his desire to be trumped by the way things actually are, or it could be that the seeking of this object or state (take happiness or virtue for example) is desired even in the case where the goal cannot

be realized. In the second case it is the pursuit itself that is the ultimate *res* that is running the motivational show.

So far I have only answered the question of how the Dominance theory deals with a non-existent object by saying that many of the cases that we take to be for a non-existent object we take in that way only because we have analyzed the object of desire according to the traditional method in which we assume a *de dicto* reading of the object in question. Thus, one answer to the question of what we do with desires for non-existent objects, as in the case of the Easter bunny, is to say that the motivating desire is not really a desire to see the Easter bunny, but a desire for happiness, fulfillment or some other such thing. The child in question might believe that he or she is motivated by a desire to meet the Easter bunny, but an observer who tries to explain the child's behavior upon arriving in the room next door (a room that we assume does not contain a living breathing Easter bunny) might make some ready hypotheses about the hierarchy of desire into which the child's supposed desire to meet the Easter bunny was placed. A child might, despite seeming very excited to see the Easter bunny beforehand, show no disappointment upon entering the Easter bunnyless room. A child might even show relief—meeting the Easter bunny was an exciting proposition, but one that produced more anxiety than could be easily handled. The child might be satisfied if the room is filled with chocolate Easter bunnies, or toy Easter bunnies or pet bunnies dressed in Easter suits. Even if the child plunges into the depths of disappointment and despair, an observation about what sort of rationalizations calm her will also be revealing to an observer who is trying to figure out how the Easter bunny fits into her overarching goals. An observer might conclude that it makes sense to think that the child desires to meet the Easter bunny, or that it doesn't—but only once observations about how such a meeting seemed to fit into this overall hierarchy have been made. Of course here we must once again recall that, due to the fact that we are making use of behaviorist criteria, we might not be able to distinguish the object of desire from an object concerning which the child is considerably deluded. The more deluded an agent is about what is actually the case—the more false beliefs she has—the harder it will be for *anyone* (agent or observer) to uncover the actual object to which her desire is oriented.<sup>16</sup>

But the Dominance theory has more to say than this. Sometimes, as detailed above, a desire that seems to have a non-existent object doesn't actually. But the other alternative is that the desire in question is incoherent. That is, the Dominance theory can interpret desires for supposedly non-existent objects as incoherent desires.

Let us be more specific about how our desires are actualized based on the hierarchy of desire outlined in the Dominance theory. All desires for actions and objects operate within the context of a hierarchy of desire, where there is an overall desire for the good that directs our desires for particular things. This context is our desire for the best end available to us in the situation we are in. Penner proposes that we understand desires to have a means-ends structure that operates according to a general formula that contains a substitution clause. This general formula is: I have the desire to do *whatever it is that will be of the most benefit to me in my present*

*circumstances*. Our beliefs about the particulars of our current situation, and our further beliefs that take the form of predictions about what might happen as the result of the various possible actions that we could perform, dictate a very particularized reformulation of the substitution clause (which I have placed in italics above). This *general* formula for desire is not *itself* the sort of thing that can motivate behavior. We act only once we have substituted a *particular* action in the appropriate way. This substitution of a particular action—one that involves a specific object and takes place at a specific time (“I wish to drive in this particular lane, here, to my right, right now”)—results in what Penner calls an “executive desire.” An executive desire integrates our beliefs into our hierarchical motivational structure yielding the motivation to perform a particular action. When we offer an explanation for a piece of purposeful behavior, we must cite an executive desire.

Using this means-ends hierarchy (captured by the formula that results in the executive desire) as a framework, we can make a distinction between cases of performing those actions that are the best means to our happiness and those that are not by making a distinction between “doing what we want” and “doing what seems best to us.”<sup>17</sup> The distinction works by illuminating two different scenarios that might unfold when we try to do the action that is the best means to the best end in our particular, current situation. When we do what we want, we *desire* that which is *in fact* the best end available to us in the situation we are in, and we *do* the action that is *in fact* the best means to that best available end. In contrast, when we do what seems best to us, we *still* desire that which is *in fact* the best means to the best end available to us in the situation we are in, and we *think* that the action we perform is the best means available to that end. However, in this case, it turns out that the action we perform is *not in fact* the best means available to that best available end.

The way that the Dominance theory will deal with desires that are not for something that is an actual means toward our happiness (because they, for example, do not exist) is to place the actions that stem from them into the category of doing what seems best. In this way, the theory acknowledges that our executive desires orient us toward a real world—a world that has the final say over whether or not our desire is satisfied. A world that is comprised by a reality that is independent of any of the ways that we think about or represent it to ourselves.

The Dominance theory allows the actual act that we perform (which is not aligned with our happiness) and the actual good that we desire (whether we know what it is or not) to work together in a specific way in order to explain our action. The theory does this by noting that we always intend to seek whatever is actually good—whatever is actually best for us in the situation at hand, and whatever is the actual best means to that best end. No one ever pursues any object or does any action on the hypothesis that it is (only) *apparently* good. The Dominance theory will use this intention to allow the way the world really is, and what is actually good for us, to dominate our desire in the formation of our executive desire. As a result, our intentions will integrate the actually best action with the action that the agent has determined is actually best and has plugged into the substitution clause to form the relevant executive desire.

This is the aspect of desire that Penner believes is analogous to the mode of reference that Donnellan (1966) spoke of in elucidating a new way of thinking about Russell's notion of definite descriptions. While Donnellan discussed reference outside of psychological contexts, Penner thinks that what Socrates thought about reference within psychological contexts works with the intentions of the agent in the same way. Penner and Rowe (1994) reason that the success of this reference is due to the fact that a certain feature of the agent's psychological state underlies the act of referring. When we refer to an object, we do so with the intention that any details we get wrong in our description of the object to which we want to refer will be overridden by whatever is actually true about the object.

In the martini glass example, Jones intends for his reference to overcome any discrepancy between how he represents things to himself and how they are. Thus, the Dominance theory adds a sort of "escape clause" to the general formula that results in an executive desire: I want whatever means to whatever ends is best for me in my current situation *and* I want my reference to what is best for me to overcome any discrepancy between how I represent things to myself and how they are.

It makes sense to think that this sort of reference underlies our investigation of any hypothesis, whether about the causes of behavior (as in the case of desire) or otherwise. If a biblical historian wants to investigate the truth of the claim that Jonah was swallowed by a big fish, it can't be the case that the historian's reference to Jonah is mediated only by her belief that Jonah was swallowed by a big fish. That is, it can't be the case that the historian only succeeds in referring to Jonah if her belief that he was swallowed by a big fish is true (that her reference only succeeds under the description "man who was swallowed by a big fish"). If that were the case, then the historian would be asking whether or not "the person who was swallowed by a big fish was swallowed by a big fish"—a question that can't fail to be answered affirmatively.

This insight carries implications for how we should understand a desire for something that is wanted as a means to a further end. In our driver example, we hypothesized that the driver wanted to drive in the right lane. We hypothesized this on the basis that he moved into the right lane. But the driver's lane change was premised on his belief that it was the fastest lane. That is, the driver wanted (inside) however things actually were with respect to the lane into which he switched (outside) to dominate his representation of that action as the action that he wanted to do. He wanted to change into the fastest lane, *whichever action it was*, but he mistook changing into the right lane for the action he actually wanted to perform. Now, the driver's executive desire explains why he switched into the right lane without our having to attribute a desire for switching into the right lane to him. The driver's false beliefs have allowed him to formulate an incoherent executive desire. It is a desire to do whatever is best—in this case, to drive in the fastest lane—but it is also a desire that orients him to do the actual action that he did (switching into the slower lane). Further, these incoherent elements are integrated by his desire that any discrepancies between the action that is really best for him and the action that he does be overridden in favor of what is actually best for him. But the integration does not, in this case, result in a coherent desire. The desire that *causes* him to drive in the right,



slower lane is not an unqualified desire to do so. It is a desire to do what is best for him (whatever that is) that has been disoriented by false beliefs. It is a desire for the good that has “gone bad.”

The driver’s desire is still a desire for the actual good, even though it oriented him to the slower lane, for two reasons: first, it is only by assuming that he desired to do the action that was the best means to the best end available to him that we can ever account for how the action that he actually did (switching into the slower lane) was plugged into the substitution clause. Second, it is a further feature of his desire for the good—his intention that any discrepancy between the action actually best for him and the action that he actually did be overridden in favor of what was actually best for him—that allows us to understand his desire as one that referred to a real, physical, action—namely, the switching into a faster lane. So, the executive desire to do a particular best action is, at the same time, the executive desire that explains why he switched into the slower lane.

Thus, the formulation of desire in accordance with the Dominance theory of desire details how we can explain why someone switched into the slower lane even when accomplishing this lane change was not the dominant object of his desire. Once we understand all of the features of his executive desire, we can insist that the driver has a desire to switch into the left lane, but is oriented in a way that explains his changing into the right lane. That is, the driver has substituted “get home as fast as possible,” and then “drive in the fastest lane” and then “change into the right lane,” in sequence for the *whatever* clause, forming an executive desire. The full expression of the general structure of an executive desire is, “*I want whatever means to whatever end is the best for me in my current situation, and I also want my reference to what is best for me to overcome any discrepancy between how I represent things to myself and how they are.*” It is this that has become particularized in each case where we purposefully perform an action.

The “and” in the italicized statement above shows that there are actually two processes of integration of insides and outsides that take place when an executive desire is formed. To illustrate the two moments at which integration of outsides and insides is afforded by the Dominance theory, let us return to the case of the six year old who goes to see the Easter bunny and is devastated and inconsolable upon finding that the room does not contain a living, breathing Easter bunny. So inconsolable is she that she shuns all things Easter for the rest of her life and is always suspicious of adult-sponsored holiday events for children.<sup>18</sup> It is her intention, as captured by the escape clause, which will result in the two coordinated moments at which insides are dominated by outsides in the formation of each executive desire.<sup>19</sup> For one, the way an action will actually turn out is allowed to dominate the way we conceive of the action when we are about to perform it. In the case of the child, her conception of meeting the Easter bunny as an act which it is possible to perform would be overridden so that when she intends to perform the action, she is referring to the action she actually will perform. This is not to say that she knows which action she will perform—none of her misconceptions have been cleared away—it is just that now, when she refers to the action she has in mind to execute, she refers to the action she is actually about to perform. It is important that she is referring to the actual action

that she will carry out, as it is that action (whatever she is about to do, with all of its properties, whatever they are, known and unknown) which needs to be involved in the second bit of overriding as governed by her intention to do what actually fits into her hierarchy of desire—what actually furthers her happiness. The second bit of overriding involves her intention to have her—possibly misconceived—desire to do that actual action be overridden in favor of whatever action is *actually* best for her. This enables outsides to triumph over insides, when it comes to determining what she wishes. This overriding is desired and intended in the event that the way the world is determines that the action that one will actually perform is not the best available means to the best available end for one, in one's present circumstances. In the child's case, this would mean that her desire to perform the action of "meeting a living breathing Easter bunny" would be overridden in such a way that it is not the case that she wishes to perform the action that she will actually do (the action of walking into the room next door whatever is in it and whatever that entails, known and unknown). Instead, this action is not desired; it only seems best to her. For she desires to do the best action available to her (with all of its properties, known and unknown) and that is not what she will do. In summary, one bit of overriding fixes the reference of the action to the actual action to be performed in the interest of fixing desire on the good; the other bit fixes desire on the good.

Although the child doesn't have a full-blooded desire to meet the Easter bunny in our example (meeting the Easter bunny only "seems best" to her), her desire for the good explains why she went to meet the Easter bunny willingly. It explains why the Easter bunny—however she represented it to herself—is featured in her executive desire. The escape clause provided by the Dominance theory brings outsides to bear on the content of our intentional states. Contrary to what is dictated by the traditional view, it is not, here, the case that a desire for an end produces *ceteris paribus* a desire for the means *thought* necessary to bring it about. Rather, our desire for an end (our own good in the present circumstances) produces *ceteris paribus* a desire for the means *actually necessary* to bring it about. At the same time, it also invokes our beliefs, some of which may be false, in orienting us to a particular action toward a specific object at an exact time. This will also result in an executive desire that is—in an impure way—*for* an action that only seemed best to us. While I will often have false beliefs about what I desire and about what is good for me, and while these beliefs might be cited in explanations for why I behave the way I do, the way an object appears to me cannot simply be cited in describing the object of my desire.

I have argued that both a theory which interprets the object of desire in a *de re* fashion and the Dominance theory, which brings intentions (insides) to bear on objects in the real world (outsides), are more successful at explaining behavior in experimental contexts than is the traditional sort of theory which restricts its attention to the *dictum* or the "insides" when trying to understand the content of a desire. I have argued this on the basis that the first two types of theory explain *more* behavior—they explain both the individual action in question and any actions that are performed as a consequence of that action.

How are we to understand the superior explanatory power of theories which allow real objects in the world to play a role in desire attribution? What are the

shortcomings of the traditional view that separates insides from outsides and pays attention only to the insides? Proponents of the traditional view attempt to compensate for the fact that a person's executive desire sometimes harbors an incoherence that results in a disparity between the action she does (with all of its properties, known and unknown) and the one she intended to do (with all of its properties, known and unknown). Frege's purpose in distinguishing sense from reference was to render our executive desires coherent. But this cannot always be done, as sometimes an executive desire will be the product of inconsistent beliefs; Frege, therefore, stipulates his solution at a price. Distinguishing sense from reference and understanding an executive desire to make use only of sense in establishing its object makes it impossible for an individual to interact *psychologically* with the actual (as opposed to the apparent) world. That is, the action that we actually do is often not the action that we conceive of ourselves as performing. Frege equates the intended action with the action that we conceive of ourselves as performing. He thinks our desire relates us to this action, an action that only has the properties that we conceive of it as having. But often, the way the world is prevents the way we conceive of the action from being a real possibility. So, we are stuck interacting with our conception of our action to the exclusion of interacting with any possible—or soon to be actual—action. Our desires don't put us in touch with the actions that we actually execute, or with the world in which they come about.

The traditional view treats desire as a two-place relation between a subject and an object. When we focus on experimental contexts, we begin to understand that desire is actually a three-place relation between subject, object and *a further end that the subject hopes to achieve*. This is also why experimental contexts allow us to see that the assumption that subjects know what they desire is problematic. Subjects readily grant that they want the object upon which they have fixated for the purpose of bringing about a further desired end. This seems to render subjects' statements about the objects of their desires sufficiently provisional so as to allow their discovery that a certain object does not bring about their desired end to override their claim that they did indeed desire it.

Let's go back to our driver example. The traditional reading of this example tries to render the driver's beliefs coherent, and so does not allow us to make substitutions in psychological contexts *salva veritate*. A theory that separates insides from outsides would force us to say the following:

The driver desired the lane into which he changed.

He changed to the right lane.

However, the driver did not desire the right lane.

In order to allow the driver to have a coherent desire, we are encouraged to understand the verb "desire" in a way that makes it the case that the driver can both:

- (1) have wanted the right lane without wanting the slower lane, even though they are identical;
- (2) have wanted both the right lane and the faster lane, even though they are mutually exclusive.

Notice that given (1) and (2) there is no actual, *physical* object that can be construed as what the driver wanted. There is no possible action that can be the actual action he wished to perform. The solution, on the standard treatment, is to say that the driver desired an apparent object:

The driver desired the lane that appeared to him to be the fastest.<sup>20</sup>

But this object is either the right lane or something other than a physical object. Let's argue against the first alternative—that it is the right lane. It is completely intuitive to think that when the driver moved into the right lane he was not satisfied that he did what he wanted to do. Let's vindicate this by showing that it is what observers (who must evaluate his behavior from the outside) would conclude.

Observing the driver, and seeing him move into the right lane, we might make the following hypothesis about how to explain his behavior:

The driver wanted to drive in the right lane.

Having made our hypothesis, we now continue our observations to see if we can confirm our hypothesis. If we are right about the right lane being what the driver wanted, his moving into the right lane should make him stop seeking whatever it was that he wanted. However, suppose that upon entering the right lane, the driver continues to check his mirrors and seem agitated—he continues to exhibit “lane-changing-behavior.”

Did the driver want to drive in the right lane? The fact that he continued to exhibit lane-changing-behavior seems to count against that hypothesis. But, it also doesn't seem to make sense that in keeping with our (1) and (2) he wanted something other than an actual action or object. Yet, this is the awkward position in which the traditional Fregean theory leaves us; the object of the psychological verb is the sense and not the reference of the term used. He wanted an apparent object or his sense of the word “fastest lane.” A theory which renders our desires coherent at the price of making us able to interact psychologically only with something like a Fregean sense is not a useful tool for explaining our behavior.

## 8.7 Conclusion

Penner's Dominance theory vindicates most of what I concluded in my argument for *de re* attribution and, so, all of the claims a *de dicto* theorist might argue should not be countenanced. The Dominance theory shows that, in order to explain behavior, we must assume: (1) that desires have a means-ends structure, (2) that explanatory desire attributions should not be read in a *de dicto* fashion because the *res* (the way the world or object actually is) plays a significant role in desire attribution, (3) that we should not assume that looking at how people represent the object of their desire to themselves should trump how those objects actually are when we attribute

desires to them—we should not assume that agents know what they want. In vindicating these claims the Dominance theory also vindicates my proposal that we should always make desire attributions the way we would in experimental contexts where we have no a priori biases concerning what is motivating a subject's behavior. What the Dominance theory of desire does not vindicate is the naive assumption that if *de dicto* attributions don't work then it must be the case that *de re* attributions do.

## Notes

1. Frege (1892/1952) first developed this explanation.
2. In fact, I have already made this challenge in my (1996).
3. Davidson (1980, pp. 4, 9–19); Dretske (1988, pp. 81, 109–15).
4. Of course it must be *actual* (not *apparent*) termination. We must also describe the behavior to be explained in a non-tautologous manner. See Reshotko (1996, pp. 158–60) for a complete explanation of these qualifications.
5. Or she wanted more food than she got. See Reshotko (1996, pp. 159–60).
6. See note 11, below.
7. This is the view that it makes sense to attribute to the Socrates of Plato's Group I dialogues. For this reason, I take the thesis of this article to be central to our understanding of the Socratic desire for the good.
8. I argue that this is a less plausible picture of the hierarchy of desire in my (2006, chs. 2, 3).
9. Of course one might ask whether the ultimate goal should be understood *de re* or *de dicto*. I will answer this question when I talk about desires for non-existent objects. See also note 10, below.
10. Practical contexts often demand that we artificially treat the subject's description of his or her goal as definitive. In doing so, we regard something that is actually a proximate goal, motivated by a proximate desire, as an ultimate goal motivated by an ultimate desire. Our driver might ultimately desire his own happiness, but we treat his desire to get home as fast as possible as his ultimate goal (taking his description of it to be definitive in this case), we are forced to give *de re* attributions of all the desires that are proximate to that desire in this context. See Reshotko (1996, pp. 168–9).
11. Note that the objection is made on the basis of the scope of the Objector's preferred hypothesis,

But now her argument has no explanatory value because it cannot explain Oedipus' "acting as if his desire has been satisfied behavior." (Huston 2000, p. 188)

12. Baker (1982) gives arguments designed to show that neither *de re* nor *de dicto* desire attributions can explain behavior (p. 380). She also attempts to develop a theory that incorporates both propositional attitudes and factual elements in order to develop appropriate explanations. Her attempt to tie down the object and subject of explanatory claims by referencing them indexically (pp. 381–387) seems to stand in for the type of reference that Donnellan's example indicates. Still she concedes that it is not at all clear that a coherent explanation will result. It seems that at least part of the problem is her attempt to preserve the incorrigibility of a subject with respect to his or her beliefs and desires.
13. It seems Santas himself (2003, pp. 99–100) has this intuition about how far the dominance theory of desire should be extended and how it might be limited.
14. Penner argues for this and labels it "maxhap" in his (2005). Happiness also seems to be identified with maxhap at *Prt.* 358c-d.
15. An argument for this Parmenidean account of what it is to exist would take us too far afield. It does seem to me, however, that the reasonable position to take is that anything to which we

- must refer in order to explain behavior—and which cannot be reduced to the appearances of an agent—has to be something and, therefore, exist.
16. As discussed more thoroughly in the last paragraph of Section 8.3.
  17. This terminology comes from Plato's *Grg.* 467a–468e.
  18. I say this in order to ensure that we are dealing with a case of a child whose desire to meet the Easter bunny is not easily explained away as a desire that is easily satisfied by—and is therefore *for*—some other existent object.
  19. It is tempting to think of them as not only coordinated, but sequential: first the reference of the action needs to be fixed on the one which will actually be performed, and then the action which is actually good needs to dominate the action which is going to be performed. But, really the two work from the ends toward the middle, so to speak. We start out with a desire to have everything about our desire align itself with what is actually good and the work to get this alignment to happen proceeds—with varying degrees of success—from there.
  20. Or, some will want to say, “the right lane under the description the ‘fastest lane’.”

## Bibliography

- Baker, L.R. 1982. “*De Re* Belief in Action.” *Philosophical Review* XCI(3):363–87.
- Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.
- Donnellan, K. 1966. “Reference and Definite Descriptions.” *Philosophical Review* 75(3):281–304.
- Dretske, F. 1988. *Explaining Behavior*. Cambridge: MIT Press.
- Frege, G. 1892/1952. “On Sense and Reference.” In *Translations from the Philosophical Writings of Gottlob Frege*, edited by P. Geach and M. Black. Oxford: Basil Blackwell.
- Huston, M. 2000. “Defending *De Dicto*.” *Ratio* XIII(2):186–90.
- Kripke, S. 1972. *Naming and Necessity*. Cambridge: Harvard University Press.
- Penner, T. 2005. “Socratic Ethics: Ultra-Realism, Determinism and Ethical Truth.” In *Norms, Virtue, and Objectivity: Issues in Ancient and Modern Ethics*, edited by C. Gill, 157–88. Oxford: Oxford University Press.
- Penner, T., and C.J. Rowe. 1994. “The Desire for the Good: Is the *Meno* Inconsistent with the *Gorgias*?” *Phronesis* 39(1):1–25.
- Reshotko, N. 1996. “Do Explanatory Desire Attributions Generate Opaque Contexts?” *Ratio*, IX(2):153–70.
- Reshotko, N. 2006. *Socratic Virtue: Making the Best of the Neither-Good-nor-Bad*. Cambridge: Cambridge University Press.
- Santas, G.X. 2003. “Penner, Plato, and Sachs on Justice and Happiness.” In *Desire, Identity, and Existence*, edited by N. Reshotko, 95–107. Edmonton: Academic Printing and Publishing.
- Stampe, D. 1987. “The Authority of Desire.” *Philosophical Review* 96:335–81.
- Stich, S. 1983. *From Folk Psychology to Cognitive Science*. Cambridge: MIT Press.

# Chapter 9

## The Good and the Just in Plato's *Gorgias*

Christopher Rowe

### 9.1 Background

What has become the traditional Anglophone view of Plato's writing divides it up into three periods: "early," "middle," and "late." "Early" usually means "Socratic," i.e., closer to the thought of the historical Socrates; "middle" tends to mean "including reference to a theory of 'separated' Forms" (*vel sim.*); "late" means anything after that. (The "late" dialogues, on this traditional, Anglophone view, are a collection of dialogues that have rather little in common, except that the kind of philosophy they represent seems—to those who wish to see it that way—closer to what we moderns, or we modern Anglophones, call "philosophy."<sup>1</sup>) Nowadays, however, this way of looking at the dialogues—let us call it the "developmentalist" view—looks distinctly less attractive than it once did, notwithstanding the support that such a reading appears to derive from Aristotle's reading of Plato (that is, given the emphasis it places on that—fundamentally Aristotelian—point about "separation"). The main reason for the decline in popularity of the "developmentalist" reading is the recognition that the developmental model has little or nothing to support it *apart* from Aristotle—and a basic psychological plausibility: what more

---

The present paper has already appeared, in a more primitive form, in Barbarić (2005, pp. 73–92); it is presented here with the permission of the publishers. The piece was, in part (as its title is meant to suggest), originally stimulated by Jerry's *Goodness and Justice*—a book that changed my own thinking on Socratic/Platonic ethics when it first came out, and one to which I continually return. The present paper is dedicated to the author of *Goodness and Justice* (and of much good in the study of Greek philosophy) with gratitude and affection. The paper is the second in a series of three papers on the *Gorgias*, all of them sharing a virtually identical first section ("Background"), and an overlapping second ("The problem of the *Gorgias*"). The first paper in the series, "A problem in the *Gorgias*: how is punishment supposed to help with intellectual error?", together with a small part of the third, has appeared in C. Bobonich and P. Destrée (2007, pp. 19–40), and the third, 'The Moral Psychology of the *Gorgias*', in M. Erler and L. Brisson (2007, pp. 90–101). The three papers are now consolidated as [chapter 4](#) of Rowe (2007c).

C. Rowe (✉)  
Durham University, Durham, UK  
e-mail: c.j.rowe@dur.ac.uk



plausible, so the argument goes, and more natural, than to suppose that Plato started by reproducing, or exploring, what was essentially his master Socrates' thinking, but then moved on, beyond Socrates (especially in metaphysics, if one takes Aristotle's line)—and finally entered a period of mature reflection, in which, perhaps, he abandoned some of the optimistic constructions of his "middle" period?<sup>2</sup> For if we take, just by itself, the evidence afforded by the measurement of Plato's style,<sup>3</sup> what we seem to find is an early group which contains *both* "Socratic" dialogues, i.e., dialogues untouched by "middle-period" Form-theory, *and* three of the central dialogues that contain that very theory: *Cratylus*, *Phaedo* and *Symposium*.<sup>4</sup> We may, of course, choose to ignore this plain fact, and carry on as normal; but it should at least be unsettling, for those of us who have tended to rely on the traditional early-middle-late division, to discover that, for all we know, Plato may have been writing "middle-period" dialogues even while he was writing "early" ones.<sup>5</sup>

My own inference from the situation as I have described it is that a re-think is needed. But in any case my collaboration with Terry Penner, and especially our work on the *Lysis*,<sup>6</sup> has convinced me that the real division among those dialogues not labeled as "late"—"late" dialogues I leave to one side, in the present context—is to be made in relation to a different theory: not the "theory of Forms" (whatever we decide that theory is, and whatever we think "separation" is),<sup>7</sup> but rather a particular theory, which Aristotle recognizes as Socrates',<sup>8</sup> about *human motivation*: the theory commonly labeled as "intellectualism," although the precise nature of Socratic intellectualism is frequently misstated and misunderstood.<sup>9</sup> The *Lysis* turns out to be a pretty single-minded statement, and exploration, of the Socratic intellectualist position; and the consequence is that that position can no longer be written off<sup>10</sup> as an isolated feature, limited to a controversial argument—based on a variety of hedonism—that Socrates introduces against Protagoras at *Protagoras* 351e ff.<sup>11</sup> Once properly understood (especially with the help of the *Lysis*), intellectualism is revealed as key to the proper appreciation of the argumentation of a range of dialogues that includes the *Symposium* as well as the group of dialogues traditionally labeled as "Socratic." Yet in Book IV of the *Republic* Socrates seems specifically to *reject* intellectualism,<sup>12</sup> and numerous other dialogues clearly imply its rejection. At the same time, whatever interpretation we put on the Platonic theory of Forms, i.e., as "separated" or otherwise, *that* theory seems to have rather few implications for any part of what Socrates either was about, historically, or appears to be about in any of those dialogues that it may be appropriate to label as "Socratic."<sup>13</sup> Plato's thinking about Forms, or in general his thinking about metaphysics and epistemology, *by itself* tends merely to add to, and does not significantly change, the ideas that he inherited from Socrates.<sup>14</sup>

Given all of this, the dialogues in question<sup>15</sup> will still tend naturally to fall into two groups—not, now, by the Aristotelian (metaphysical) criterion, but rather according to whether they (1) presuppose, explore, or otherwise make use of, or alternatively (2) reject or ignore this (apparently) Socratic theory of human

motivation. *The* turning-point in Plato, both in terms of his relationship to Socrates and, I propose, in general,<sup>16</sup> is marked by that moment when he ceases to be interested in, and indeed positively begins to argue against, that theory.<sup>17</sup> If it is true that there are “intellectualist” dialogues, on the one hand, and “non-intellectualist” (or “anti-intellectualist”) dialogues on the other, the easiest hypothesis seems to be that Plato began by thinking the Socratic position powerful, and central (for in numerous dialogues it *is* central), but later came to think differently, and to suppose that he needed a different line, one that would improve on, make good what he had come to see as the defects of, the original Socratic account of human action. Or, at any rate, so I myself hypothesize.

What is this “intellectualist” theory of motivation (or, perhaps better, theory of *action*; it is not just a theory of desire)? Briefly, and at bottom, it consists in the claims (1) that all human agents always and only desire the good; (2) that what they desire is the *real* good, not the *apparent* good; and (3) that what we do on any occasion is determined by this desire together with whatever beliefs we have about what will in fact contribute to our real good. Hence the label “intellectualist”: we only ever do what we *think* will be good for us. So “virtue [or ‘excellence’] is knowledge,” or would be if it could ever be realized, and also “is one”—because, if the theory is correct, and is nevertheless to make room for virtues/excellences like justice, courage, and the rest, then they must all be a matter of making the right calculations in relation to good and bad. (“Virtue is knowledge,” then, in that it is a matter of knowledge of what is truly good and truly bad; and it is one for the same reason.) And given all of this, it will simply be impossible for anyone to do, or (as I prefer to put it) go, wrong willingly; one can only go wrong through ignorance.

*This* is what the Socrates of the *Republic* then famously denies: that is, when he argues in Book IV for the existence of two irrational parts of the soul, which can—and this is the crucial point—actually overcome reason, perhaps even knowledge. The argument in *Republic* may indeed be taken as going out of its way to underline the conflict between its conclusion and the “intellectualist” position.<sup>18</sup> And the difference is quite fundamental. For if we all possess irrational elements or parts that are capable of causing us to act independently of, or even in direct contravention of, what our reason tells us to do, then it will plainly be insufficient merely to *talk* to people, in the way that the Socrates of the dialogues seems to do, in order to change their behavior; we shall need to deal with their irrational parts as well—which will require irrational, i.e., *political* and *rhetorical*, means. It is no accident, I propose, that a large part of the rest of the *Republic* is occupied with talk about political institutions, including a state-run education system involving what is in many respects a kind of conditioning.<sup>19</sup> How different *this* Socrates is from the essentially a-political, or un-political, Socrates of the *Apology*, or the *Crito*, or . . . That other Socrates claimed that what was needed was philosophy, dialectic, thinking things through. But now that is no longer enough: one may think as much as one likes, and yet if we pay them no heed, our irrational elements may still ambush us, by night if not by day.<sup>20</sup>

## 9.2 The Problem of the *Gorgias*

So the proposal is that the so-called “early” and “middle” dialogues (that is, again, all apart from the late dialogues) would be better divided—roughly speaking—into pre-*Republic*, on the one hand, and *Republic* plus post-*Republic* on the other. That will, evidently, give us a new “early” and a new “middle,” but it seems better to avoid that terminology, insofar as “middle” tends to be so heavily associated with the move to the new metaphysics (“separated” Forms, etc.). In any case, my claim is that some of the relevant dialogues feature the “Socratic,” intellectualist, theory of action, and some feature a radically different, if rather more familiar, kind of theory of action. I say “more familiar”: who nowadays would accept the Socratic “denial of *akrasia*”—or, to put it better, his explanation of what others, including the Plato of the *Republic*, treat as “lack of control,” or, in that spectacular bit of English mistranslation, “weakness of will”?<sup>21</sup> We moderns are ourselves liable to take it for granted that we can be *overcome* by desire—we are all used to saying “I don’t know what came over me,” “I couldn’t help it,” and so on. “No,” says Socrates, “you are wrong—you *could* help it; nothing *made* you do it. You acted as you did because of the state of your beliefs (so, if you don’t like what you did, you’d better do something about your beliefs).” Or so he would respond in the ambit of some of the dialogues (the ones I am proposing to call truly “Socratic,” including the *Symposium*—that old “middle” dialogue, which is nonetheless thoroughly *intellectualist* in its treatment of human behavior);<sup>22</sup> in others, (perhaps) starting from the *Republic*, it looks as if he comes over more to what I have called the familiar modern position—though even then he will be rather less inclined than we often are to accept it as any sort of *defense* that “something came over me.” (“Pull yourself together!” will be his response—even while apparently still holding that such cases are, in Aristotelian terms, involuntary.<sup>23</sup> But of course, as the *Republic* shows, he thinks that some will be more capable of pulling *themselves* together than others; others will need external help.)

Now in this whole context, the *Gorgias* may well seem to be something of an anomaly.<sup>24</sup> For on the one hand the *Gorgias* contains one of the most spectacular applications of the Socratic theory of action, in the shape of Socrates’ claim that orators and tyrants *have no power*—a claim from which he not only never retreats in the rest of the dialogue, but on which he seems to build even more surprising, paradoxical, even (apparently) comical claims. Those apparently enviable people, who—so Gorgias has claimed—can do whatever they want, in fact—Socrates says—do *nothing* they want, only what seems best to them. “How ridiculous!” responds Polus. But of course Socrates is perfectly serious: they *don’t* do what they want. Why not? Because they don’t have the knowledge to enable them to distinguish properly between good and bad, and lacking that, they fail to get what is really good for them—which must be what they want; doesn’t *everyone* want what is really good for them? Whoever was satisfied with what merely *seems* good, and isn’t in fact so? This, surely, is the full Socratic position.<sup>25</sup>

Yet on the other hand—and this is what makes the dialogue seem anomalous—the *Gorgias* is likely, to most readers, to look in significant respects significantly

un-Socratic. Perhaps most strikingly of all, it appears an un-Socratically *political* dialogue,<sup>26</sup> one that in numerous respects seems to foreshadow the *Republic*: the whole discussion, after all, centers around issues of power and the place—if any—of rhetoric in society; and in one of the climactic moments of the dialogue, Socrates the philosopher declares himself, bizarrely, to be (possibly) the only true statesman in existence.<sup>27</sup> It will then probably appear entirely consonant with this strongly political aspect of the *Gorgias* that the dialogue has a great deal to say about *punishment*; for after all it is the state, or the city, that punishes. And punishment, surely, uses force, which I have argued ought strictly to be useless on a Socratic account of motivation and action. However there is something else that looks—prima facie—even more obviously un-Socratic about the *Gorgias*. For from almost the beginning of his argument with Callicles in the last third of the dialogue, Socrates relies heavily on the idea that we need to *control ourselves*, and especially our desires; and that at once seems to involve him in allowing for the possibility of our *failing* to control ourselves and our desires—or, in other words, of his allowing for the possibility of *akrasia*, and the kind of divided soul that goes with it. But how can Socrates do that, while remaining Socrates?<sup>28</sup>

The specific problem of the prominence of the theme of punishment in the *Gorgias* was the subject of the first of the series to which the present paper belongs<sup>29</sup>; while the perhaps more fundamental problem, highlighted by Terry Irwin,<sup>30</sup> of the juxtaposition of Socratic intellectualism with an emphasis on self-control was the subject more than 20 years ago of a useful treatment by John Cooper,<sup>31</sup> in which he claims to show—against Irwin—that the moral psychology of the *Gorgias* is in fact Socratic through and through. I am myself more in accord with the outcome of Cooper's treatment than with the method by which it is reached; for I find myself in basic disagreement with Cooper over what precisely Socratic intellectualism (i.e., his theory of action) is about. However Cooper's short paper provides a kind of justification for my putting off, for the present, the biggest of the issues I have raised,<sup>32</sup> and concentrating on another, and more specific, aspect of the problem of the *Gorgias* (if that is, generally stated, whether the dialogue is Socratic or un-Socratic in its outlook, or some strange mixture of the two). This more specific aspect is what the Socrates of the dialogue wants to say about *the Good*, or more generally, about the object or objects of human action; and especially about why it is good to be *just*.<sup>33</sup>

My attention will focus on Socrates' encounter with Polus in the second act of the *Gorgias*. More specifically, it will focus on Socrates' claim—one that he also says (474b) that *Polus, and everyone else*, accepts as well—that doing injustice is worse, for the agent, than would be his having it done to him.<sup>34</sup> What is of special interest here, in the context of a discussion of Socratic and (apparently) un-Socratic elements in the *Gorgias*, is the way in which Socrates is (apparently) content to argue to a conclusion that he himself evidently owns from what *Polus* accepts, and indeed accepts with enthusiasm<sup>35</sup>—and this, against the background of a general context that shows Socrates to be, or at any rate ought to put him, at some distance from the very things that *Polus* accepts and that he, Socrates, relies on for his conclusion. *Polus'* position is—or at any rate he *says* it is (see below)—that doing injustice is

better for the agent than suffering it, but more shameful (*aischion*); by having Polus agree that fineness/beauty (*kallos*, treated as the opposite of *aischos*) is a matter of what is useful (good), pleasant, or useful/good *and* pleasant, and that doing injustice is not more painful (474d ff.), Socrates compels him to change horses and say that after all it is doing injustice that is the worse thing, not suffering it. But now elsewhere in the *Gorgias* Socrates adopts a distinctly radical view of what is good for us: so, for example, at 511c–512c, when he talks about the modest claims that would be made by any expert navigator: yes, Socrates says, the navigator saves people’s lives, but who is to say—the navigator certainly won’t—that he will always have done them a service by saving them? No one else in the dialogue, and certainly not Polus, would even flirt with such a radical suggestion. How, then, can Socrates be content to argue to a conclusion about what is better on the basis of Polus’, evidently rather more everyday, notion of what counts as better? Here, one might suppose, is another case of the Socrates of the *Gorgias* carrying his convictions lightly.<sup>36</sup> Again, the Socrates of the *Lysis* and the *Symposium* seems straightforwardly to *identify* the fine/beautiful with the good<sup>37</sup>; what business has he, then, proposing to treat it as *either* the good *or* the pleasant, *or* both (if it is the same Socrates, which is what is currently at issue)?

Now it could be that we should simply say that it is all a piece of bad, or, probably worse, merely opportunistic argument. On this version of events, the end-result would be that *Polus* agrees that doing injustice is worse, according to *Socrates*’ conception of “worse,” by means of Socrates’ playing on *Polus*’ (or else, if *Polus* oughtn’t to own up to it, nobody’s) notion of good and bad. But the consequences of such a diagnosis look just too unpalatable: if Socrates is prepared to argue like that, one might reasonably ask why we should take him seriously at all.<sup>38</sup> I prefer, then, to set up the problem in the terms suggested above: why is it that—or is it really the case that—the Socrates of the *Gorgias* is so selective in applying what appear to be distinctively his own, radical, beliefs?

### 9.3 A Standard Modern Solution; and an Alternative Suggestion

So what is going on? We may begin with Socrates’ initial suggestion that *Polus* and everybody else in fact agrees with him (Socrates) that doing injustice is worse for the agent than suffering it—and this in the face of *Polus*’ emphatic and repeated assertion of the exact opposite. What Socrates has in mind here might perhaps be simply that what he says *Polus* and everyone else agree about is somehow a consequence, unforeseen by them, of what they do in fact quite openly say (i.e., that doing injustice is more shameful but better than having it done to one). After all, that is how things seem to work out. But to read Socrates in this manner is to put ourselves danger of reintroducing, in a different form, the very sort of problem I have just outlined: after all, unless we have reason for believing what *Polus* and everybody else says, we have little reason for being impressed by any alleged consequences of those beliefs, and neither has Socrates. He himself has already expressed scorn at *Polus* for relying on witness-statements for the purposes of refutation (471e–472d).

Part of what he is doing is certainly to provide elucidation of what Polus says: when, in clarifying the latter's claim that doing injustice is *aischion*, i.e., less *kalon*, than suffering it, Socrates proposes that the *kalon* is either the good or the pleasant or both, Polus agrees enthusiastically—"you're defining finely, now, Socrates, when you define the fine by pleasure and good" (475a2–4, Irwin transl.). But now what kind of move is this; why does Polus agree to it; and should he agree to it?

What is probably the most influential modern kind of answer to these questions is to say that Socrates believes, perhaps passionately, in his conclusion about the relative merits of doing and suffering injustice, and is trying to find an argument for it; it is just not a very good argument, and the proposed analysis of the *kalon*—into beneficial, pleasant, or both—is its weak point. Terry Irwin, in his commentary on the *Gorgias*, sums up how he sees the situation in the following way:

Socrates has either a valid argument with an implausible and undefended premise [sc. "x is finer (more *kalon*) than y if and only if either x is pleasanter than y or x is more beneficial than y" for the subject, i.e., the agent in action x], or a more plausible premise [sc. "x is finer (more *kalon*) than y if and only if either x is pleasanter than y or x is more beneficial than y" for those concerned, explained as "those taken to be relevant in any particular case"], and an invalid argument.<sup>39</sup>

Socrates—so Irwin proposes—just "fails to disambiguate" his claim; Polus meanwhile "accepts Socrates' way of 'defining' . . . the *kalon*, presumably because it sounds realistic and down-to-earth, referring to people's pleasure and advantage."<sup>40</sup>

So, on this account,<sup>41</sup> Polus is duped, and Socrates does not know what he is doing. According to Irwin, the analysis of *kalon* into good, pleasant or both is to be explained by linguistic usage:

Socrates' two conditions for being *kalon*, pleasure . . . and use . . . or benefit . . . are explained by the wide range of "*kalon*." "Pleasure" is meant to explain how faces, buildings, etc. [the reference here is to Socrates' *epagôgê* in support of his analysis of *kalon* in 474d–475a] are *kalon* in themselves. "Use" explains how shoes are *kalon* for walking. . . .<sup>42</sup>

This diagnosis of the argument is at best disappointing: however splendid the argument's formal conclusion, it all turns out to be a matter of the partially sighted leading the blind, and not making a great success of it. However, even as it faces us with this dismal prospect, Irwin's comment accidentally points to an escape-route. Irwin continues:

But ["use"] need not apply only to means to further ends. Plato often says that what is good is beneficial, *M[eno]* 88e, *R[epublic]* 379b, and he asks whether justice is profitable, *R[epublic]* 367c, 392b. He may only be saying that anyone who has justice is better off than anyone who lacks it; it might be an intrinsic or an instrumental good. . . . At the same time it is unwise to assume that Plato is clear in the *G[orgias]*, as he is at *R[epublic]* 357b–358a, about the different kinds of goods he includes under "beneficial"; he may regard them all as instrumental. Nor is his position on hedonism clear. . . .<sup>43</sup>

Yet—and it is here that Irwin's comment is so suggestive—surely Plato (Socrates) *has* made clear what goods he "includes" under "beneficial," if the beneficial is the good, as the general context confirms. Only a few pages further back, Socrates has put forward to Polus some rather specific proposals about the good: that it is the



good that we desire; that this good is the end, whatever it—the end—is, for which we do whatever we do; that it is not what we do that we desire, but what we do what we do *for*, and so on (467c–468d).

But in that case it is, I think, impolite as well as uncharitable not at least to begin by supposing that when the good comes back into the discussion, as it does in the context we are presently considering (“which is *worse*, do you think, Polus, doing injustice or suffering it?”, 474c5–6), those previous proposals are meant still to be in play. In other words, Socrates has a *theory* about the good, which—so I am proposing—we should expect to make itself felt, other things being equal, whenever the subject of the good comes up. Just so it will reappear at 499–500, in Socrates’ conversation with Callicles:

... for I take it we agreed that we must do everything for the sake of goods, if you remember—Polus and I. Do you [Callicles] agree with us too, that the good is the end of all actions, and that for the sake of it we should do all the other things, not do it for the sake of the other things? (499e6–500a1, Irwin transl.; Callicles does agree.)

Here as in 467–8, Socrates commits himself to the idea that it is only ends that are goods, at least in the first instance, and *pari passu* that means to those ends are not goods (at least in the same way).<sup>44</sup> The idea perhaps looks bizarre, but—since he recurs to it—it is one that the Socrates of the *Gorgias* evidently sponsors; it is not just something he proposes, out of devilment or to get his own way, and then drops. Why then should we not imagine it also to be presupposed in 474–5?

Grounds for not imagining it presupposed there might be looked for in that initial statement of Socrates’ at 474b2–4: “For I think that I *and you and other men* believe that doing injustice is worse than suffering it.” This after all suggests a common ground for the belief in question, and so—one might propose to infer—the absence of any *special* Socratic elements. But if Socrates does have a special theory of the good up his sleeve, the common ground will apparently not extend very far. Socrates and Polus “and other men” may all turn out to believe that doing injustice is worse than suffering it, because they all accept that it is *aischion*; but if Socrates’ argument relies on ordinary notions about what it is to be *aischion*, then—the old problem<sup>45</sup>—the agreement he trumpets seems hardly very solid or useful.

So we seem to be on the horns of a dilemma: either Socrates is suppressing his own beliefs in order to claim agreement, or claiming agreement while implicitly having Polus and everyone else sign up to *his*, Socrates’, account of things. There is, however, a way of avoiding either of these readings of the argument. We should notice that in 482d–483c Callicles will implicitly accept Socrates’ argument: where Polus went wrong was just in allowing that doing injustice, rather than suffering it, was *aischion*, after rightly claiming that suffering it was worse—for “everything is more shameful which is also worse.”<sup>46</sup> Callicles, at any rate, has no argument with Socrates’ analysis of *kalon* into good or pleasant or both; it is just that he has a different notion of what is to count as good. This, I propose, is where things stand between Socrates and Polus (and everyone else) too. It is not that Socrates assumes Polus will accept everything he would want to say about the fine (beautiful/ admirable?) and the good. Rather, he proposes to Polus that the fine is the good,



the pleasant, or both, whatever account one goes on to give of the good (and the pleasant); meanwhile *he* actually has a developed account of it to give, while Polus does not, though some—different—sort of notion of the good must be implied in his claim that suffering injustice is worse.

At this point, however, I suggest that we need to make a significant adjustment. In 474–5, Socrates allows Polus the option of treating the pleasant as a separate criterion of choice and action,<sup>47</sup> apart from the good. This is the main basis that Socrates offers him for the distinction, on which he—Polus—insists (474c9–d2), between the fine and the good: the fine is what is good *or* what is pleasant, or both. However later on, in his conversation with Callicles, Socrates will reject that position for himself.

“Do you agree [Callicles] . . . that the good is the end of all actions, and that for the sake of it we should do all the other things. . . .?” “I do.” “Then for the sake of goods we should do other things, including pleasant things, not good things for the sake of pleasant things?” “Quite.” “Now is it for anyone to select which kinds of pleasant things are good and which bad. . . .?”<sup>48</sup> (499e7–500a5)

In other words, as I take it, the sensible, technically correct,<sup>49</sup> *wise* thing to do is to consider, before going for something that one thinks will—and even in fact will—prove pleasant, whether it is a *good* thing to go for it or not. (“Wise,” because, if it is goods we want, what we “should,” *dei*, do is always to think whether this, now, will give us what we want.)<sup>50</sup> The question that then immediately arises is whether Socrates will want to say, as Polus seems to do, that things can be fine and not good, i.e., just because they are pleasant, and that too seems to be ruled out by the conversation with Callicles, part of which turns on the unacceptability, whether to Socrates or even to Callicles, of just such a position.<sup>51</sup> In short, left to himself Socrates would, to all appearances, have been quite content with the move “And so [doing injustice] is worse [than suffering it] too, if indeed it is *aischion*” (474c8–9)—a move which he makes elsewhere,<sup>52</sup> but which Polus emphatically rejects, so forcing him to take a different tack. But in the event no use is actually made of the possibility “fine because (merely) pleasant” in 474–5, since even on Polus’ account doing injustice is not less *pleasant* than suffering it. So in fact, if the fine is the good or the pleasant or both, and doing injustice is less fine than suffering it, it must be worse, whether on Socrates’ or on Polus’ or on anyone else’s view of what the good is. The argument is not the one Socrates would have chosen, but one that is forced on him by the state of Polus’ beliefs. At the same time, he commits himself to nothing he doesn’t believe, and nothing in the argument depends on anything he doesn’t believe—and so we escape the horns of the dilemma as outlined.

Or so it seems. However one set of questions still remains—the very ones with which the present section began. Even granted that anyone (apart from Callicles) will say that doing injustice is the less fine thing, the less admirable (let us finally opt for the latter as a translation of *kalon*, as a way of combining good and pleasant), what compulsion is there on Polus, or on us, to accept Socrates’ analysis of the *kalon*? Even if it is enough to convince Polus, the *epagôgê* in 474d3–475a2 has generally tended to strike readers as less than impressive. Apart from the sort of objection raised by Irwin (which I myself tend to think beside the point),<sup>53</sup> it perhaps

looks rather cursory; and from a modern point of view the analysis itself might well seem to come from nowhere—or, again, just too conveniently for Socrates. Granted, he makes that addition “beneficial *or pleasant*” for Polus’ sake, and Polus seems to like it well enough (when he approves of the analysis, it is pleasure he mentions first).<sup>54</sup> But why?

Modern skepticism about Socrates’ move in this instance is surely connected with the fact that, as we should put it, what he is talking about is, after all, a question of *moral choice*: the *moral*—so our feeling might be—should not, perhaps cannot, be reduced to the (merely) *prudential*. So the *epagôgê* is not merely unconvincing, but is evidence of a serious category mistake: we might even put up with Socrates’ trying it on in this way, but surely his contemporaries had a clear and distinct notion of the category of the moral? Is it not bound up with the very language of justice and the virtues? Here is a splendidly measured statement of the main issue, by Irwin:

It is doubtful whether we should speak of different *senses* of “*kalon*”, especially of a “moral sense” of the term. It is hard to say that the Greeks were aware of using it in different senses, as we might be aware of using “bank” in different senses for a riverside and a finance house; a statue and a brave action might both be *kalon*, though different properties would make them *kalon*. It is even harder to pick out a “moral sense” of the term; it is not easy to pick out specifically moral terms in Plato anyhow. . . . There is no reason to believe that the use of “*kalon*” for what we call moral properties indicates that the Greeks have a particularly “aesthetic” attitude to morality, as the translation “beautiful” might suggest. *Kalon* covers what is admirable from the aesthetic, the agricultural, the industrial, the prudential, and the moral point of view, and there is no reason to think any point of view primarily determines the sense or associations of the term. (1979, p. 154)

Part of the upshot of this comment is presumably that—notwithstanding the absence of a “moral *sense*” of the term—*kalon* cannot, on Irwin’s view, be reduced to the pleasant and the useful/beneficial without remainder, or at any rate in the context of a discussion of justice. Just as “a statue and a brave action might both be *kalon*, though different properties would make them *kalon*,” so “bodies, colors, shapes, sounds”<sup>55</sup> and a just action<sup>56</sup> will by the same reasoning both be *kalon/kala*, but by virtue of possessing different properties. But that is just what Socrates clearly denies, and what Plato has Polus (and later, by implication, Callicles) deny. Here is the issue: Socrates, and Plato to whatever extent he is here identifying with Socrates, will have their special reasons for denying it,<sup>57</sup> whereas—the objection runs—Plato will have no business foisting the same position on Polus, let alone “other men,” especially if that is supposed to include us the readers.<sup>58</sup>

In response, I point out that on the interpretation I have suggested above, “foisting” anything on anyone is precisely what Socrates/Plato is trying *not* to do. Having ascertained that Polus (emphatically) does not want to follow him<sup>59</sup> in identifying *kalon* and *ôphelimon/agathon*, Socrates gives him an alternative. So he acknowledges that Polus, and by implication everyone else, do not share his own position; and through that alternative he tries to spell out just what Polus’ (and the general) position is on the nature of the *kalon* (the admirable), and—in effect—how it differs from his own position. This, he suggests, is by virtue of supposing that the admirable can also involve *pleasure*—as it clearly will do in the case of artifacts,

“practices,” and so on, but—and here I speculate, because Socrates does not spell it out—as it will also do in cases of admirable *actions*, or in the case where we have to compare two things in the sphere of action in respect of their admirableness: e.g., doing versus suffering injustice. “Well,” I imagine Socrates as saying to Polus,

you say that doing injustice is less admirable than suffering it. As you agree, this means that you will be saying that it is either worse or less pleasant (or both). To the extent that you might even toy with the idea of thinking it less pleasant, that reflects the genuine sense of *distress* (pain) that you feel at the idea of doing injustice. But at the same time you admit that that distress is less than the distress that you think will be caused to you by having that same injustice inflicted on you. So it can't be the distress to you that would be caused by doing injustice that makes you call it less admirable; it must rather be that you think it will be *worse* for you—despite the fact that you actually claim that it's the other thing, having it done to you, that's worse.

When Socrates proposes, and Polus accepts, the addition of the pleasant to the analysis of the *kalon* (“you're defining finely, now, Socrates, when you define the fine by pleasure and good”),<sup>60</sup> the effect is to mark the difference that at least Polus feels between *kalon* and *agathon*—the very sort of difference, one might imagine, that we are likely to describe in terms of the moral, or perhaps the aesthetic.<sup>61</sup> Yet the pleasure in question is evidently not any sort of *moral* pleasure, insofar as it is immediately commensurable with grosser kinds: any distress Polus has at the prospect of doing injustice is trumped by the greater distress he will feel at being treated unjustly. That might not be a matter of physical pain; it might just be the pain or distress caused by losing face. But *moral* considerations are surely not supposed to be so easily set aside (how, otherwise, will anyone ever be brought to act unselfishly?).

Has Plato shortchanged Polus, by not allowing him to decide (what we should call) a moral choice on moral grounds? Probably not, since Polus is not obviously a person of high moral sensitivity. Has Plato short-changed his, and Polus', contemporaries? (After all, the argument is supposed to apply to everyone.) This is a more difficult question to answer. It may be that the Greeks had a relatively unformed notion of the moral; and if we think that notion important (even central), we might conclude that it was Plato's business to shape it better, not to propose abandoning it in the way that he appears to be doing.<sup>62</sup> Yet we can also see him as *challenging* justifications of the kinds of behavior we admire in terms of the moral, and replacing such justifications with something he conceives of as stronger: that behaving admirably is simply good *for us*. Callicles will allege that Polus was merely shamed into saying that doing injustice was more shameful than suffering it; it is only convention (*nomos*: Irwin prefers “rule”) that makes people say such things, whereas in fact it is what is worse—suffering injustice—that is shameful (482d–483b). Socrates pays Polus the compliment of supposing that he is saying what he actually believes; he also pays him the further compliment of suggesting that he actually believes something that he, Socrates, thinks is philosophically defensible.<sup>63</sup>

What is important to note about this is that it is not a matter of mere philosophical charity on Socrates' part. In working out exactly what it is that Polus and “the others” believe in believing that doing injustice is more shameful, less admirable,

than suffering it, he is also, given his theory of desire, pointing to what he thinks it is that drives them, in all their actions and choices: the universal desire for the good.<sup>64</sup> It is to this (the good) that their descriptions—in terms of the admirable and the shameful—really refer. To the extent that they do not realize this, Socrates no doubt overreaches himself, with a touch of mischievousness, when he proposes that they actually agree with him. “Agreeing” is hardly something one can be properly said to do without knowing about it. To that degree, Socrates is after all only inferring “agreement” to his conclusion (doing injustice is worse), on the basis of premises his interlocutor admits to, without that interlocutor’s dreaming for a moment that they lead just *there*. My claim is just that this is not all there is to it. What Socrates gets out of Polus, understood as *Socrates* understands it, is no more than he thinks he would get out of him (and anyone—perhaps even Callicles),<sup>65</sup> with enough dialectic: in Polus’ case, as the fully spelled-out version of what he is saying—intends to say—when he uses the words “doing injustice is more shameful, *aischion*, than suffering it.” But that would simply be the moment at which Polus fully understood his own, true, motivation—towards a good (for him as agent) that excludes doing injustice under any conditions.<sup>66</sup>

As for what that good is, Socrates does not say here, or indeed anywhere else in the *Gorgias*. But if the Socrates of the *Gorgias* is, as I propose, an *intellectualist* Socrates (see Section 9.1 above), then the answer to that question is bound to have something to do with *knowledge*.<sup>67</sup> For if what we all want is what is good, then everything will depend on our finding out, with as much precision as possible, what that good (for us) is—or, in other words, on our acquiring the kind of expertise possessed by the special “craftsman,” *technikos*, that Socrates and Callicles agree will be needed to sort out which pleasures are to count as goods in each sphere.<sup>68</sup>

There is much in the above that requires further thought and/or justification. The conclusion of the paper, however, given the necessary limitations on its length and ambitions, is that so far from being opportunistic, unfair to Polus, or any of the other things with which it has been or could be charged, *Gorgias* 474–5 is a serious—and indeed radical—piece of argumentation, which shows every sign of being of a piece with Socrates’ larger argument in the *Gorgias* (and beyond). Socrates is not to be underestimated, nor is his author; we need to expend every effort in trying to understand what they—character and author—are about before leaping in to identify the weaknesses and shortcomings of their argumentative strategies. To do otherwise is to risk using Plato’s texts merely to confirm our philosophical prejudices, and our sense that in general we ought to have moved on beyond the intellectual and cultural achievements of the fifth and fourth centuries BCE. That would be a wasted opportunity, if in fact these texts are able to offer us a different perspective from our own, and a different vantage-point from which to review, and recognize the extent and nature of, our assumptions.<sup>69</sup>

What is more—to go back to the original project of this paper—the *Gorgias* shows up as through and through Socratic. That is, it rests firmly on distinctly Socratic ideas. There are no compromises even in that difficult, and controversial, passage on which this paper has focused (474–5); indeed that passage, properly

understood, is a kind of restatement of core Socratic positions. Or so I have argued, and will continue to argue until shown that there is a better way to interpret the argument Plato has Socrates present.

## Notes

1. For a recent restatement of this traditional view of the dialogues as dividing into early-middle-late, see Fine (2003, n. 1 to Introduction). Fine refers back, for a defense of the traditional view, to Vlastos (1991, chs. 2, 3); but these two chapters are mostly concerned with a different proposal ("that through a 'Socrates' in Plato we can come to know the thought of the Socrates of history": Vlastos 1991, p. 81), and presuppose the traditional division of Plato's works rather than defending it.
2. Such a picture of the evolution of Plato's thought is likely to appear particularly appealing against the background of a general assumption that progress in philosophy is linear, and of the more particular assumption that Aristotle is a much more evolved specimen of a philosopher than his teacher Plato, and Plato than *his* teacher, Socrates. Fine's book (2003) reflects both assumptions, which are indeed endemic among British and American scholars. I myself regard such assumptions as at least unhelpful, to the extent that they interfere with our giving Plato, and Socrates, a decent hearing; and the present essay firmly rejects them. That is to say, I am not in the least inclined to treat the kinds of positions I shall attribute to the Socrates of the *Gorgias* (who is, in my present view, not so distantly related to the real Socrates) as quaint, or simply false. Part of the point of the present attempt to recover what this Socrates is saying is that in my view—which I share with my friend, colleague, and co-author Terry Penner—it stands a rather good chance of being *true*.
3. This is not to say that we must necessarily believe everything we are told by the stylometrists, whose track record—at least in more recent times—has not been uniformly good. However (1) at least some of their conclusions appear to be reasonably firm; and (2) in any case the traditional early-middle-late paradigm has generally been thought (mistakenly: see below) to be supported by those firmer conclusions.
4. See especially Kahn (1996, 2002):

At first sight, the division into three stylistic groups [proposed by a number of scholars working mainly in the nineteenth century] seems to confirm [the] theory of Plato's development [in question], since all of his "Socratic" dialogues are firmly located in the earliest group. But this first sight is misleading. The central group does not at all coincide with what are called the "middle" dialogues, since the intermediate group defined stylistically includes both *Parmenides* and *Theaetetus*, which are generally counted as "late" from a developmental point of view; on the other hand, the "early" group includes *Symposium*, *Phaedo*, and *Cratylus*. A traditional developmentalist who recognizes that the stylistic division is chronological must simply accept the fact that Plato's stylistic and philosophical developments do not proceed at the same pace. (2002, p. 96)

5. Which is merely a different way of saying what Kahn says in the last sentence cited in the preceding footnote.
6. See Penner and Rowe (2005).
7. "...[Aristotle] writes as though separation is the big differentiator between Plato and Socrates," says Gail Fine (2003, p. 298). She thinks this untrue; "commitment to separation ['capacity for independent existence': pp. 255–6] is as muted in the middle dialogues as lack of commitment to it is in the Socratic dialogues." "Separation is not, however, the only feature Aristotle points to in differentiating Plato from Socrates; and perhaps other of his claims are on firmer ground. Aristotle also claims, for example, that for Socrates, unlike Plato, all universals are sensible, that is, are sensible properties. Now Plato, as we have seen, accepts NR

[non-reducibility]; Forms are nonsensible properties, properties irreducible to, and indefinable in terms of, sensible properties” (p. 298). It is metaphysics, then, that still seems to divide Plato from Socrates, for Fine.

8. And which he seems to regard simply as false, and so uninteresting, and/or a mere historical relic. See, e.g., *Nicomachean Ethics* III.4, where the theory is dismissed as self-contradictory: “the consequence, for those who say that the object of wish is the good, is that what the person making an incorrect choice wishes for is not wished for (for if it is wished for, it will also be good; but in fact it may have been bad)” (1113a17–19; how can something that is wished for—as it will be on Aristotle’s account—also *not* be wished for?). Plato’s mistake about universals (as Aristotle conceives it) is, by contrast, interesting and important. For Aristotle’s recognition of the theory dismissed in *NE* III.4 as *Socratic*, see, e.g., Penner (2003) and Rowe (2003a).
9. For one splendidly clear statement of the general outline of the theory in question, see Taylor (2000, pp. 62–3). This is, I suppose, what Brickhouse and Smith (2002) have called—somewhat puzzlingly: see the last sentence of this note—“the traditional account of Socratic intellectualism” (p. 22). Brickhouse and Smith “attribute to Socrates a more complex moral psychology, one that retains a central tenet of ‘pure intellectualism,’ namely, that no one acts contrary to what he or she believes is best, but which also assigns a specific causal role to non-rational desires” (ibid.)—a role that will require reason to *control* them. If this were indeed Socrates’ view, then—I suggest—it will not merely be that “Plato’s mature moral psychology owes a greater debt to its Socratic predecessor than most commentators have realized” (p. 35); Socrates’ moral psychology will be virtually indistinguishable from that of the *Republic* (cf. section II below). A specific criticism that should be made of the Brickhouse-Smith paper—which of course bears directly on the issues discussed in the present paper—is that it allows a *myth* to determine central elements in Socratic thinking. For what I myself propose to make of talk of “incurables” in the myth of the *Gorgias*, see Rowe (2007a, n. 42)—and for (what I take to be) other and not dissimilar misstatements of the essentials of Socratic intellectualism, see Cooper (1982) and Irwin (1979).
10. As it is, for example, by Kahn (1996, ch. 8).
11. Or, alternatively and more generally, dismissed as unworthy of a good philosopher like Plato. For a slightly more extended treatment of the issues here, see Rowe (2003b).
12. That is, in the course of arguing for the existence of three parts of the soul, one rational and two irrational, the irrational parts (respectively “spirited” and “appetitive”) themselves being capable of causing the agent to act *even contrary to reason*. Such actions are ruled out by the “intellectualist” model, according to which all desires are for the (real) good, and the only difference between agents who get things wrong and those who get things right is in the state of their beliefs (see below). (In fact, in Rowe 2007c, I come to argue *against* the view that Socrates/Plato rejects intellectualism, either in *Republic* IV or anywhere else: he finds a—surprising—way of retaining it despite the introduction of irrational parts. That, however, is a battle that it would be uneconomical to attempt to fight in the context of the present paper.)
13. I.e., either by the traditional criterion (i.e., showing no evidence of “middle period” metaphysics) or by the criterion I am here proposing (i.e., whether resting on or alternatively rejecting intellectualist premises).
14. *Pace* Fine, e.g., in the chapter of her (2003) cited in note 7 above.
15. Once again, for the purposes of the present argument I continue to restrict myself to those dialogues traditionally labeled “early” and “middle.”
16. The question of what motivates us human beings is, I presume, likely to be central on anyone’s account of Plato’s philosophy; my own view is that it is, and remains, closer to the center of Plato’s thinking than anything in the spheres of metaphysics and ontology, or of epistemology, though I recognize that I may well be in a minority in holding this.
17. It is of course theoretically possible that Plato alternated: now using/applying the one sort of theory, now the other. The consequences of the two theories are, however, so large (see Rowe 2003b, p. 28 ff.) that I count this as no more than a theoretical possibility.



18. At *Republic* IV, 438a–439b Socrates argues specifically that there are desires (“appetites”) that are not good-directed (cf. n. 13 above): “Therefore, let no one catch us unprepared or disturb us by claiming that no one has an appetite for drink but rather good drink, nor food but good food, on the grounds that everyone after all has appetite for [‘desires’: *epithumei*] good things, so that if thirst is an appetite, it will be an appetite for good drink. . . .” (Grube/Reeve transl., in Cooper 1997). (It is this interpretation of the *Republic* IV argument that I propose to reject in Rowe 2007c: see n. 13 above.)
19. Again, see Rowe (2003b).
20. See *Republic* IX, 571b4–572a1 (cited, in Grube/Reeve transl., with omissions):

“Some of our unnecessary pleasures and desires seem to me to be lawless. They are probably [are likely to be: *kinduneuousi*] present in everyone, but they are held in check by the laws and by the better desires in alliance with reason. In a few people, they have been eliminated entirely or only a few weak ones remain, while in others they are stronger and more numerous.” “What desires do you mean?” “Those that are awakened in sleep, when the rest of the soul—the rational, gentle, and ruling part—slumbers. Then the beastly and savage part, full of drink, casts off sleep and seeks to find a way to gratify itself. . . . On the other hand, I suppose that someone who is healthy and moderate with himself goes to sleep only after having done the following: First, he rouses his rational part and feasts it on fine arguments and speculations; second, he neither starves nor feasts his appetites, so that they will slumber and not disturb his best part with either pleasure or their pain. . . .”

21. Mistranslation, because it presupposes either that the Greeks had a concept of the will, or that any true picture of the world must inevitably make room for such a concept. Both presuppositions are questionable, to the extent that a concept of the will surfaced only centuries later, to provide for the resolution of mental conflicts—conflicts, that is, of just the sort whose existence Socrates, and others (notably the Stoics), deny.
22. So that, strikingly, passionate or romantic love, *erôs*, can be described (by Socrates and the seer Diotima) without any recourse to the concept of irrational, non-good-directed desires.
23. Just so Socrates’ counterpart as main speaker in the *Laws* is still to be found insisting, Socratically, that “no one does/goes wrong willingly” (*Laws* V 731c [no one commits justice willingly], 734b [everyone is *akolastos* unwillingly]; IX 860d [all bad people are bad, with respect to everything, unwillingly]). It is what is *really* good that at least some part even of the Platonic divided soul still desires.
24. Vlastos (1991, ch. 2) treats the *Gorgias* as straightforwardly one of “the dialogues of Plato’s earlier period” (p. 46); evidently he misses the kinds of problems that I here identify—problems that suggest at least some kind of *transitional* status for the *Gorgias*. For Vlastos, “transitional” dialogues are early ones that are merely missing the “elenchus” according to his unnecessarily narrow notion of “elenchus” (i.e., “examination,” “challenge,” “[attempt at] refutation,” which actually appears to be a standard part of Plato’s notion of philosophical method: see, e.g., Penner and Rowe 2005). Fine does treat the *Gorgias* as “transitional” (2003, p. 1), but she does not state her grounds for doing so. From the perspective of the present series of papers, however, the most important reference will be to Irwin’s commentary on the *Gorgias* (Irwin 1979), which sees the dialogue as using, and failing to reconcile, two different approaches to “good-independent” desires: “(1) The unhealthy soul has a faulty conception of its good, and needs to be restrained because otherwise its desires—all good-dependent—will mislead it. (2) Its strong good-independent desires make it incontinent [‘weak-willed’], so that it needs control. . . . The conclusions of [the] two lines of argument [depending on these different approaches] in the dialogue are never satisfactorily reconciled” (p. 218). What the present paper, and the third in the series to which it belongs, i.e., Rowe (2007a), set out to resist is something very like Irwin’s account here; though I differ significantly in the way I state (1), the Socratic position. See following note.



25. See Penner (1991). One absolutely crucial difference between Penner's interpretation and Irwin's (see preceding note) of Socrates' position is that Penner sees it as insisting—however paradoxically—that we *only* desire what is *really* good for us. Insofar as Irwin talks of ["good-dependent"] desires as potentially "*misleading*" the soul, and so apparently being responsible for its "faulty conception of its good" (passage cited in preceding n.), he evidently does not take this line. ("Good-dependent," then, will have a distinctly weaker force than in Penner's interpretation.) My own interpretation follows Penner's and not Irwin's.
26. The Socrates of the *Gorgias*, as one of Vlastos' "dialogues of Plato's earlier period," ought to lack that "elaborate political theory [sc. of the *Republic*] whose ranking order of constitutions places democracy with the worst of contemporary forms of government, lower than timocracy and oligarchy, preferable only to lawless tyranny" (Vlastos 1991, p. 46). That, I suppose, he does lack; yet in political terms the *Gorgias* goes far beyond the *Crito*, which Vlastos seems to take as defining the political dimension of the "early" dialogues—not least in virtue of that stunning moment, at (*Gorgias*) 521–522 (to which I shall shortly advert in the main text, and in Section 9.3) when Socrates claims to be—perhaps—the only true statesman alive. It is surely less far from here to the philosopher-ruler of the *Republic* than it is to citizen Socrates in the *Crito*.
27. See preceding note. (For my own reading of *Gorgias* 521–522, see Rowe 2007a.)
28. See the preceding two paragraphs. The denial of the possibility of conflict between reason and passion (desire) seems to be the hallmark of the Socratic position: what people want, what they are passionate about, is their real good, and their real good only (which is why tyrants and orators have no power).
29. I.e., in Rowe (2007a).
30. See note 24 above.
31. Cooper (1982); see also Cooper (1999).
32. On which see Rowe (2007b).
33. Here is another way, and perhaps the most obvious of all, in which the *Gorgias* is likely to remind us of the *Republic*—that it attempts to provide a reasoned argument for our behaving justly towards others (which is, after all, the chief underlying theme of at least nine books of that longer work).
34. Gregory Vlastos, in "Does Socrates cheat?" (1991, ch. 5), spells out the claim differently: "*exactly what* [Socrates] *means*" is that "[d]oing wrong is worse *for the wrongdoer* than is suffering it *for the victim*" (pp. 144–5). But Socrates himself, as Vlastos points out on his own account, simply states the claim as "doing wrong is worse than having it done to one" (so, e.g., at 474b), and there seems to me to be no basis in the text for supposing that the doer and the sufferer are meant to be different people—indeed, the text actually tells the other way: see, e.g., 469b, and esp. 474b7 (Polus to Socrates) "would *you* prefer to be treated unjustly than to do injustice?" In one way there is no real disagreement here; the issue just is: should one prefer, e.g., to be the one torturing, or the one tortured? But Vlastos' introduction of different perspectives appears to be what allows him to accuse Socrates of a fairly uninteresting sort of fallacy. I shall consider below a more generous variant on Vlastos' interpretation of the argument.
35. 475a2–4 (on which, see Section 9.3).
36. Cf., e.g., note 24 above, on the *Gorgias* as "transitional."
37. See especially *Lysis* 216c–217a, *Symposium* 204d–e.
38. True, some scholars—e.g., Kahn (1983)—are willing to accept a Socrates who merely sets out to fool his interlocutors; but I, for one, would side with Vlastos in preferring a Socrates who merely went wrong to one who cheats. (In this case, I think Socrates actually does neither.) Could Plato really be imagined as making a discussion of how to live our lives—the ultimate subject of the dialogue—hang on a mere sleight of hand, when his own preferred choice, and that of his character Socrates, includes *philosophy*?
39. Irwin (1979), on *Gorgias* 475c.
40. Irwin (1979) on 475a.

41. Irwin's general treatment here closely follows Vlastos', but is more circumspect (or "generous," the term I used earlier).
42. Irwin (1979) on *Gorgias* 474d.
43. Ibid.
44. For more on this idea, see Penner and Rowe (2005).
45. See Section 9.2.
46. This is one salvageable part of the apparently corrupt sentence uttered by Callicles at 483a7–8; see Irwin's note.
47. This, again (or so I take it), is what the passage is about—the *choice* between *doing* injustice and having it done to one.
48. I prefer "bad" here to Irwin's "evil," for *kakon*.
49. A reference to 500a5–6: "Or does it need a craftsman—a *technikos*—for each thing?"
50. Given all of this, I fail fully to understand why Irwin claims (comment on 474d: see above) that "[Socrates'] position on hedonism is [not] clear." It may not be clear at 474d, but by the time we have assembled all the available information on what Socrates' views are (i.e., not much later than 499–500), we surely know enough to say that he is not a hedonist. But my own grasp on Irwin's point here may be less than complete, and in any case it is for the most part not directly relevant to my argument.
51. 494c–e (itching and scratching; the life of the catamite).
52. See note 37, with text to note, above.
53. The objection is, broadly, that Socrates fails to say *to whom* the "fine" is allegedly beneficial and/or pleasant, and so leaves his claim fatally ambiguous (see text to n. 39 above). Socrates does not specify in this way simply (or so I think) because he takes it for granted that the benefit/pleasure in question is benefit/pleasure accorded by, or expected from, whatever thing is in question to the person judging that thing to be fine. If Socrates happens to mention "observers," in the plural, in the case of admirable (beautiful) things that accord pleasure (474d8–9), that is just a way of stating a general truth: if any individual declares a thing visually admirable (beautiful), that is because it pleases *him*, or *her*, not because it pleases someone else. It may be that it is, for us, a "far more plausible account of something's being fine or beautiful" that "anyone—onlooker or not, [sic] truly judges that *x* is beautiful if *x* gives pleasure to someone (in this case, the onlooker)" (Irwin 1979, on 475a). But if, for Socrates, the beautiful is the good, and the good is the good for us, then presumably what is beautiful is what is beautiful for us.
54. 475a2–4 again: "you're defining finely...when you define the fine by pleasure and good."
55. Some of Socrates' examples in 474d.
56. Let us say: choosing to suffer injustice rather than to do it.
57. I.e., whatever reasons lie behind the typical Socratic identification of *kalon* and *agathon* (broadly, that ultimately all our actions, even those we perform "for others," are motivated by desire for our *own* good).
58. However different Greek assumptions may have been from ours, it is hard to read much, say, of Greek tragedy without supposing that the characters at least broadly share our sense of admiration, shame, or revulsion at broadly the same sorts of actions.
59. I.e., again, on my reconstruction of what Socrates' own position is: see above ("Apparently you don't think the same thing is fine and good..." 474c9–d2).
60. Again, we should notice that Polus mentions pleasure first; as I suggest, it is the addition of pleasure to the good that he approves of.
61. Cf. the passage last quoted from Irwin (1979) above.
62. I.e., by reducing the admirable to the good (for the agent), and the shameful to the bad (for the agent).
63. As I do, and as Terry Penner does: see Penner and Rowe (2005).
64. See Sections 9.1 and 9.2 above.
65. He would, of course, be starting further back with Callicles, in that Callicles even denies that doing injustice is *aischion* (a different dialectical tack would be needed—and in fact Socrates

- will approach him rather differently, in the final movement of the dialogue). There are signs, however, that Socrates is not prepared to give up even on him; and after all (on Socrates' theory) Callicles wants the same as everyone else (the real good). See Rowe (2007b).
66. For a longer, more fully argued, and more precise treatment of the very complex issues involved here, see Part II of Penner and Rowe (2005). In case of any apparent or indeed actual (but unintended) disagreement between the treatment here and that in *Plato's Lysis*, the latter is to be preferred.
  67. See also Rowe (2007a, b).
  68. Thus there will, no doubt, be pleasures to be had from behaving justly, and behaving well in general. Socrates is not bound to agree with Polus that suffering injustice is really less pleasant than doing it; he might just not want to take that particular argument on. Alternatively, and reasonably, he might refuse to take on the extreme position that the pleasures of justice outweigh the worst pains, the worst physical suffering.
  69. I renew my thanks to Jerry Santas, who carries some of the responsibility, but none of the blame, for the argument I now hand over to him.

## Bibliography

- Annas, J., and C.J. Rowe, eds. 2002. *New Perspectives on Plato, Modern and Ancient*. Cambridge, MA: Harvard University Press.
- Barbarić, D., ed. 2005. *Platon über das Gute und die Gerechtigkeit*. Würzburg: Königshausen & Neumann.
- Bobonich, C., and P. Destrée, eds. 2007. *Akrasia in Greek Philosophy from Socrates to Plotinus (Philosophia Antiqua, vol. 106)*. Leiden: Brill.
- Brickhouse, T., and N.D. Smith. 2002. "Incurable Souls in Socratic Psychology." *Ancient Philosophy* 22:21–36.
- Cooper, J. 1982. "The *Gorgias* and Irwin's Socrates." *Review of Metaphysics* 35:577–87.
- Cooper, J., ed. 1997. *Plato: Complete Works*. Indianapolis: Hackett.
- Cooper, J. 1999. "Socrates and Plato in Plato's *Gorgias*." In *Reason and Emotion: Essays on Ancient Moral Psychology and Ethical Theory*, edited by J. Cooper, 29–75. Princeton: Princeton University Press.
- Erler, M., and L. Brisson, eds. 2007. *Gorgias-Menon: Selected Papers from the Seventh Symposium Platonicum*. Sankt Augustin: Academia.
- Fine, G. 2003. *Plato on Knowledge and Forms: Selected Essays*. Oxford: Clarendon Press.
- Irwin, T., ed., trans. 1979. *Plato: Gorgias*. Oxford: Clarendon Press.
- Kahn, C.H. 1983. "Drama and Dialectic in Plato's *Gorgias*." *Oxford Studies in Ancient Philosophy* 1:75–122.
- Kahn, C.H. 1996. *Plato and the Socratic Dialogue: The Philosophical Use of a Literary Form*. Cambridge: Cambridge University Press.
- Kahn, C.H. 2002. "On Platonic Chronology." In edited by J. Annas and C.J. Rowe, 93–127.
- Penner, T. 1991. "Desire and power in Socrates: The argument of *Gorgias* 466a–468e that orators and tyrants have no power in the City." *Apeiron* 24:147–202.
- Penner, T. 2003. "The historical Socrates and Plato's early dialogues: some philosophical questions." In edited by J. Annas and C.J. Rowe, 2002, 189–212.
- Penner, T., and C. Rowe. 2005. *Plato's Lysis*. Cambridge: Cambridge University Press.
- Rowe, C. 2003a. "Comments on Penner." In edited by J. Annas and C.J. Rowe, 2002, 213–25.
- Rowe, C. 2003b. "Plato, Socrates, and Developmentalism." In *Desire, Identity and Existence: Essays in Honour of T. M. Penner*, edited by N. Reshotko. Kelowna, BC: Academic Printing & Publishing.
- Rowe, C. 2007a. "A Problem in the *Gorgias*: How is Punishment Supposed to Help with Intellectual Error?" In edited by C. Bobonich and P. Destrée, 19–40.
- Rowe, C. 2007b. "The Moral Psychology of the *Gorgias*." In edited by M. Erler and L. Brisson, 90–101.

- Rowe, C. 2007c. *Plato and the Art of Philosophical Writing*. Cambridge: Cambridge University Press.
- Taylor, C.C.W. 2000. *A Very Short Introduction to Socrates*. Oxford: Clarendon Press.
- Vlastos, G. 1991. *Socrates, Ironist and Moral Philosopher*. Cambridge: Cambridge University Press.

# Chapter 10

## Socrates, Wisdom and Pedagogy

George Rudebusch

### 10.1 Introduction

Intellectualism about human virtue is the thesis that virtue is knowledge. Virtue intellectualists may be eliminative or reductive. If eliminative, they will eliminate our conventional vocabulary of virtue words—“virtue,” “piety,” “courage,” etc.—and speak only of knowledge or wisdom. If reductive, they will continue to use the conventional virtue words but understand each of them as denoting nothing but a kind of knowledge (as opposed to, say, a capacity of some other part of the soul than the intellect, such as the will or the appetites). Virtue intellectualists may be pluralists or monists. If pluralist, they identify the virtues with distinct kinds of knowledge. If monist, they identify all the virtues with one and the same kind of knowledge.<sup>1</sup> In a number of dialogues—including the *Euthyphro*, *Apology*, *Charmides*, *Euthydemus*, *Laches*, *Lysis*, *Protagoras*, and *Republic I*—Socrates gives arguments that support Reductive Monist Intellectualism (RMI) about human virtue.

Socrates’ arguments make RMI both an attractive philosophical hypothesis in its own right and an attractive interpretation of Socrates’ own theory of piety, courage, and the other virtues. On this interpretation, expressions such as “piety,” “the knowledge how to serve the gods,” “courage,” “the knowledge what to dread and what to dare,” all refer to one and the same object, namely, the knowledge of the human good. I note that, for example, this sort of identification of piety with the whole of virtue is endorsed by Kant (at least when he sets aside non-rational religious revelation), and seems a consequence of some teachings of Hebrew prophets and Jesus.<sup>2</sup> My aim in this chapter is not to defend such profound philosophical hypotheses about piety or courage, although I shall briefly indicate the relevant Socratic arguments below. Nor is my aim here to establish RMI as the correct interpretation of Socrates. I seek here only to disarm an influential objection to RMI as the correct interpretation of Socrates.

---

G. Rudebusch (✉)  
Northern Arizona University, Flagstaff, AZ, USA  
e-mail: George.Rudebusch@nau.edu

According to this objection, RMI cannot be the correct interpretation of Socrates, because in the *Laches* Socrates seems to believe that piety, courage, etc. are but parts of virtue.

- PW/L To Laches, he says, “I would not have us begin with the whole, . . . let us begin with a part [of virtue, courage]” (190c8–10).
- PW/N To Nicias, he says, “You yourself said that courage was a part, and there were many other parts, all of which taken together are called virtue. . . . In that case, do you say the same as I? I apply the term to courage, temperance, righteousness, and the like. Would you not say the same?” (198a–b).

Likewise Socrates in the *Meno* and *Euthyphro* raises the question and uses analogies in such a way that his interlocutors, at least, take him to believe the conventional view that piety, courage, etc. are but parts of virtue.

- (PW/M) When Meno says, “Justice is virtue,” Socrates asks him, “Is it virtue or a virtue?” Then Socrates explains how roundness is but a part of shape, suggesting to Meno that Socrates sees justice in the same way with respect to virtue. Meno says in reply, “You’re right, for I also say (*kai egô legô*) that justice is but one part among others of virtue” (*Meno* 73e). Socrates repeats this part/whole agreement at 78d.
- (PW/E) When Euthyphro agrees that all that is holy is just, Socrates asks him, “And is all justice holy, too?” Then Socrates explains how shame is but a part of fear, suggesting by the analogy that piety is but a part of justice. Euthyphro replies that piety is but a part of virtue, “For you [Socrates] appear to speak rightly (*phainê(i) gar moi orthôs legein*)” (*Euth.* 12d).<sup>3</sup>

If we take Socrates’ part/whole claims and suggestions about virtue as statements of his own moral theory, we must either abandon monism as an interpretation of Socrates’ moral theory and opt for pluralism or else give up the attempt to find a coherent Socratic moral theory.<sup>4</sup>

My reply to this objection is to supplement the RMI interpretation of Socrates’ moral theory with an account of Socrates’ pedagogical technique, by showing that (Section 10.3) Socrates distinguishes three levels of attainment of wisdom; that (Section 10.4) Socrates’ life’s work is pedagogy, namely, moving people from the lowest level to the middle level of attainment of wisdom, by testing and examination; that (Section 10.5) one of Socrates’ pedagogical techniques is to test his interlocutor’s knowledge of a subject by giving him a false lead, that is, by making a misleading suggestion; that (Section 10.6) this technique is illustrated with the slave boy in the *Meno*; with the consequence that (Section 10.7) the RMI interpretation with this account of Socratic pedagogy explains as false leads Socrates’ claims to Laches and Nicias and his suggestions to Meno and Euthyphro that courage is but a part of virtue (namely, PW/L, PW/N, PW/M, and PW/E). In this way I disarm the

objection to the RMI interpretation. I begin (Section 10.2) by reviewing the case for the RMI interpretation.

## 10.2 Socratic Reductive Monist Intellectualism

In Book IV of the *Republic* there is an account of justice as a harmony of rational, semi-rational, and non-rational elements within a tripartite soul (a tripartite soul also appears in the *Phaedrus* and *Timaeus*). Book IV's non-intellectualist account of virtue is incompatible with Book I, in which Socrates argues that justice must be, precisely, wisdom (349b–350d). Other dialogues—including the *Apology*, *Charmides*, *Euthydemus*, *Laches*, *Lysis*, *Meno*, and *Protagoras*—portray a character like the Socrates of *Republic* I who gives arguments that reduce virtue to wisdom.

In the *Laches* and *Protagoras* Socrates provides a simple and compelling argument for the reduction of all goodness to wisdom: any non-wise trait which we might be tempted to identify as something good, such as confidence (*Prt.* 349e–350c) or endurance (*La.* 192b–d) will in some circumstances be foolish and bad. But neither foolish endurance nor foolish confidence is good. Hence any non-wise quality, such as confidence or endurance, is no more good than bad. And anything that is agreed to be good by the same reasoning must be wisdom. This argument justifies Socrates' surprising identification of good luck (*eutuchia*, *Euthd.* 279d), personal beauty (*kalos*, *Prt.* 309c), and “every [good] property” (*panta chrêmata*, *Prt.* 361b) with wisdom, and it entails, *pace* Protagoras (*Prt.* 350d), that even physical strength is wisdom—if strength is something good. In the protreptic of the *Euthydemus* Socrates uses a similar argument to the same conclusion: any object (apart from wisdom) that we might be tempted to identify as something good, such as wealth, health, good looks, good birth, power, or honor (*Euthd.* 279a–b), produces no benefit without wisdom and is positively bad when led by ignorance (*Euthd.* 281b–d). Depending on their use, these objects are no more good than bad; hence only wisdom is good and only ignorance bad. In the *Charmides* Socrates again identifies, on the basis of the same sort of right-use considerations, this good wisdom as the particular “knowledge of good and bad” (*tês [epistêmês] peri to agathon te kai kakon*, 174c) “whose function is human benefit” (*hês ergon estin to ôphelein hêmas*, 174d). In the protreptic of the *Lysis* he argues, again from considerations of utility, that wisdom is the only thing human beings love with desire (210a–d: we are loved only insofar as we are wise). In *Republic* IV (349a–354a) he gives a different sort of argument, from the structural similarity of craft knowledge to justice (*dikaïosunê*) and from the function of the human soul, in order to identify justice with the knowledge of human benefit (350d). The main objection to intellectualism, of course, is not its necessity to virtue but its sufficiency, for it gives no role to emotions, will, and desire in human virtue. However, in the *Protagoras* (352a–357e) Socrates effectively replies to this objection and shows that knowledge alone is sufficient for virtue by arguing that it is impossible for emotions to overpower knowledge of



the good, and in the *Meno* (77b–78b) by arguing that all humans are alike in willing and desiring what is good. These arguments identifying virtue with one type of knowledge enable us to understand Socrates' claims in the *Apology* that “virtue produces wealth and all other goods” (*Ap.* 30b) and that “nothing can harm a good man either in life or after death” (*Ap.* 41d).<sup>5</sup>

The arguments of the preceding paragraph are evidence that Socrates is a reductive intellectualist about virtue. But they do not tell us if he is a pluralist (making piety, courage, etc. distinct parts of virtue) or a monist (for whom the different names “piety,” “courage,” etc. all refer to the very same kind of knowledge). Socrates' refutation of Nicias indicates he is a monist. Nicias, like Socrates, is a reductive intellectualist: he says he has often heard an “excellent saying” from Socrates, that “every man is good in that in which he is wise” (*La.* 194c–d), and he agrees with Socrates that

(RI) “If a man knows all good and evil, . . . he lacks no virtue” (*La.* 198d).

Nicias is certainly a pluralist: he holds the conventional view that

(CP) Courage is but “a part of virtue” (*La.* 198a).

But Nicias runs into trouble when he tries to specify what part of the knowledge of good and evil courage is:

(CF) Courage is “knowledge of future good and evil” (*La.* 198c).

For, given CF, Socrates refutes Nicias by pointing out that knowledge is universal in time:

(UK) If a kind of knowledge knows a thing, it will know it at any time, past, present, or future (*La.* 198d).

Hence:

(CA) Courage, “instead of being only a part of virtue, will be all of virtue” (*La.* 199e).

Socrates, as shown in the previous paragraph, accepts reductive intellectualism (RI). The universality of any given kind of knowledge (UK) is undeniable. This argument therefore refutes the pluralist doctrine that courage is but a part of virtue (CP).<sup>6</sup>

The same sort of considerations of the identity conditions for any given kind of knowledge, such as Socrates provides (*Ion* 531e–532a, 537c–540e; *H. Mi.* 367c–369a; *Grg.* 449d–454a; *Rep.* I 332c–334a), lead to RMI as a general account of the virtues and as the best interpretation of Socrates.<sup>7</sup>

The above arguments, which reduce an apparent multiplicity of goods and virtues to one single thing, the knowledge of human advantage, allow us to explain

Socrates' reaction to Protagoras' Great Speech. Socrates is surprised that Protagoras would speak of virtue as a whole *composed of parts*.

You [Protagoras] say that virtue is teachable (*didakton*), and if I believe anyone [about this point], I believe you. But your speech surprised me. . . . For. . . at many points in your speech you spoke as if justice and temperance and piety and all these things, *taken together*, are one thing, virtue. Please go through this point in your argument more precisely for me. Is virtue one thing and justice, temperance, and piety parts of it, or are all these words I just mentioned names of one and the same thing? (*panta onomata tou autou henos ontos*, *Prt.* 329b–d)

With this question Socrates is establishing the outlines of his public debate with Protagoras, which will last for the remainder of the dialogue. Protagoras says the answer is obvious: justice, temperance, etc. “are parts of a single thing, virtue” (329d). Socrates, as the monist intellectualist interpretation is able to explain, is surprised that a wise man would hold any part/whole thesis about virtue, indicating that his own position identifies each of these alleged parts with one and the same teachable thing, hence one type of knowledge.<sup>8</sup>

### 10.3 Three Levels to Wisdom

The case for the RMI interpretation is not yet established. For we have seen passages contradictory to RMI (namely, PW/L, PW/N, PW/M, and PW/E). To explain these passages I turn now from Socrates' arguments in moral theory to the pedagogy he uses in his discussions with such pretenders to knowledge as Laches, Nicias, Meno, and Euthyphro. This pedagogy is based upon an account of attainment of wisdom.

In the *Apology* Socrates distinguishes three levels of attainment of wisdom. The highest level is “real wisdom” (*tô(i) onti sophos*, 23a), which is the property of God. The middle level is being “wisest among men” (*humôn, ô anthrôpoi, sophôtatos*, 23b), which is the property of anyone who, like Socrates, “knows that he does not possess real wisdom of any value” (*egnôken hoti oudenos axios esti tê(i) alêtheia(i) pros sophian*, 23b). The lowest level is “not being wise, but seeming wise, especially to oneself” (*dokein men einai sophos. . . malista heautô(i), einai d'ou*, 21c).

Socrates also calls this sort of wisdom *intelligence* (*phronêseôs*, 29e), and identifies it as the perfection of the soul (*tês psuchês hopôs hôs beltistê*, 29e), and as the human excellence which is righteousness (*Rep.* I 350d). Recognizing the supreme value of wisdom and accepting the oracle's statement that such wisdom is not possessed by human beings (23a), Socrates values and recommends only one sort of activity for a person at the middle level: “Each day to make arguments, engage in dialogue, and carefully examine virtue and related topics—this is the very best thing for a human being. Indeed, a life without this activity is not worth living for a human being” (38a). Socrates, to judge from these claims in the *Apology*, reduces virtue and value to wisdom and distinguishes three levels of attainment of wisdom.

Socrates presupposes these three levels in other dialogues. Meno's slave boy began at the lowest level, thinking he knew what he did not know, but after Socrates'

questioning, he reaches the middle level and “no longer thinks he knows what he does not know” (*hōsper ouk oiden, oud’ oietai eidenai, Meno* 84b) and as a result the boy “is in a better position concerning the thing he does not know” (*beltion echei peri to pragma ho ouk ê(i)dei, 84b*), because while at the lowest level he would not have “attempted to seek for or to learn what he did not know but thought he did” (*epicheirêsai zêtein ê manthanein touto ho ô(i)eto eidenai ouk eidôs, 84c*). Now at the middle level, the boy desires wisdom and in that sense has become a philosopher.<sup>9</sup>

The *Lysis* draws a three-level distinction between the Good, the Neither-good-nor-bad, and the Bad. It is only at the middle level, the Neither-good-nor-bad, that there is desire for wisdom (*philosophein, 218a*). At the highest level, just as the good body possesses health, the good souls “whether divine or human” (*eite theoi eite anthrōpoi, 218a*) possess wisdom and hence do not desire it. At the lowest level, the bad souls are so ignorant they do not even desire wisdom (*oud’ au ekeinous philosophein tous houtôs agnoian echontas hōste kakous einai; kakon gar kai amathê oudena philosophein, 218a*). What distinguishes souls at the middle level is that, “although possessing ignorance, which is bad, they are not yet so foolish and ignorant [as the lowest level], for [at the middle level] they understand that they do not know what they do not know” (*hoi echontes men to kakon touto, tēn agnoian, mēpō de hup’ autou ontes agnōmones mēde amatheis, all’ eti hēgoumenoi mē eidenai ha mē isasin, 218a–b*). In the course of the *Lysis*, we see Socrates help the boys, Lysis and Menexenus, ascend from the lowest level to the middle level.

In the *Protagoras*, too, Socrates distinguishes three levels in his interpretation of Simonides’ poem.<sup>10</sup> He calls the highest level “being (not becoming) good” (*ou. . . genesthai esthlon, . . . alla to emmenai, 340c*). As Socrates interprets the poem, “a god alone can have this privilege” (*theos an monos tout’ exoi geras, 341e*) of being good; “to be a good man is impossible [for mortals] and superhuman” (*einai andra agathon. . . adunaton kai ouk anthrōpeion, 344c*). The level of becoming (not being) good is the difficult one (*agathon. . . genesthai chalepon eiê, 340c, 344c*), and he describes this condition as “the middle” (*ta mesa, mesos, 346d*). The lowest level is “being bad” (*kakon emmenai, 344c*).

Socrates in his interpretation of Simonides makes two additional points that are consistent with the account of the three levels in the *Apology*, *Lysis*, and *Meno*. The first addition is that the middle level is good in the sense that “whatever is not bad is good” (*panta toi kala, toisi t’ aischra mē memeiktai, 346c*), meaning that the middle ground is “accepted not blamed” (*ta mesa apodechetai hōste mē psegein, 346d*). The second addition explains the first: A soul in the middle state does nothing bad (*meden kakon poiê(i), 346d*). This second point, that the philosopher does nothing bad, explains what Socrates says of himself in the *Apology*: “I am persuaded that I have done nothing unjust to anyone” (*pepeismenos dê egô mêdena adikein, 37b; cf. 37a5*). If philosophers do not perform their moral duties, they are not blameworthy, since the error is due to ignorance and their ignorance, unlike the ignorance of non-philosophers at the lowest level, cannot be due to negligence.

## 10.4 Pedagogy from Lowest to Middle Level

According to Socrates, then, there are three levels of attainment of wisdom. Only gods are at the highest level, possessing wisdom. Most human beings are at the lowest level, ignorant to such an extent that they don't even desire to find wisdom. And there are some, like Socrates, at the middle level, who are aware of their ignorance and thus desire wisdom. Obviously, people at the middle level with respect to any type of knowledge—geometry, medicine, or the human good—are incapable of knowledgeable acts. As a doctor benefits bodies by making their souls healthy, so an expert at human benefit, that is, the sage or virtuous person, can benefit human beings by making their souls wise. Just as someone who wants to be but is not yet a doctor is incapable either of healing others or of making others into doctors, so likewise Socrates, who wants to be a sage, is incapable of healing the souls of others, which is precisely to make them experts at the science of human goodness. And, because this same science of human goodness is at once piety and justice, *only the sage*, not the one who wants to be a sage, is able to do acts of beneficence, piety and justice.

Nevertheless, at the middle level, above most people at the lower level, Socrates is capable of a pedagogy that is in a sense beneficence, service to the gods, and activity on behalf of justice, as he explains. To be at the lowest level, ignorant but seeming wise, is an error damaging enough to negate any value in (“to hide from sight”) the wisdom of any other craft (*kai autôn hê plêmmeleia ekeinên tèn sophian apokruptein*, 22d–e) one may possess.<sup>11</sup> Socrates attempts on a case-by-case basis to move his fellow citizens out of this lowest level into the middle level by combining “exhortation” with “demonstration” (*philosophôn kai humin parakeleuomenos te kai endeiknumenos*, 29d), “questioning and examining and testing” each who consents (*erêsomai auton kai exetasô kai elegxô*, 29e).<sup>12</sup> This activity of Socrates, although not the true beneficence of the sage (who has the power to move people to the highest level), nonetheless is beneficence in the sense that it is as great a good as any that the Athenians have ever received (*ouden pô humin meizon agathon genesthai en têt(i) polei*, 30a). This very same activity of Socrates, although not true piety (which would perfect human beings in goodness by moving them to the highest level), nonetheless is piety—“Socrates’ service to God” (*tên emên têt(i) theôt(i) hupêresian*, 30a) in obedience to what God commands (*tauta gar keleuei ho theos*, 30a)—in the sense that Socrates is changing us from bad to not-bad as we reach the middle level. And this same activity of Socrates, although not true justice, which would know how to make each of us just, nonetheless is in a sense justice, for it is “really acting on behalf of justice” (*ton têt(i) onti machoumenon huper tou dikaiou*, 32a) by taking us out of our bad condition. Socrates’ life’s work, therefore, is inexpertly done and not strictly speaking beneficent, pious, or just, but in the relaxed sense of the *Protagoras*—“Call everything good which is not bad!” (346a)—they are good and blameless acts, and in terms of the *Apology* they are good in the sense that there is nothing better, more pious, or more just that Socrates can do.

## 10.5 False-Lead Pedagogy

Socrates' pedagogy, as displayed in all the Socratic dialogues, consists among other things in testing (*Ap.* 29e), and Nicias confirms this is his practice (*La.* 188a–b).<sup>13</sup> Sometimes the testing is a matter of asking the other for a justification or explanation of his actions:

To test Hippocrates' grit [*apopeirômenos tou Hippokratous tês rhômês*], I began examining him with a few questions. "Tell me, Hippocrates," I said, "In your present design of going to Protagoras and paying him money as a fee for his services to yourself, to whom do you consider you are resorting, and what is it that you are to become?" (*Prt.* 311b)

Socrates in his discussion with Hippocrates never misleads him with a false suggestion, nor does he need to drive him to self-contradiction to take him from the lowest level (seeming to himself to know what a sophist is, 312c: "I think I know") to the middle level (aware of his ignorance as to what a sophist is, 313c: "You appear not to know what a sophist is. . . —That is likely, Socrates, from what you have said"; *ton de sophistên hoti pot' estin phainê(i) agnoôn... eoiken, ephê, ô Sôkrates, ex hôn su legeis*). Socrates also describes himself as tested when he meets someone with radically different (and presumably false) views, as when he meets Callicles in the *Gorgias* (*basanizousin*, 486d; same verb at 487a and 487e) and must try to produce that person as a witness for the truth of his own view (as he promised to do with Polus at 474a).

The sort of test that I need for my reply in defense of the RMI interpretation of Socrates occurs when one person makes a suggestion that is false but seems true to the ignorant. Socrates describes the following cases of such testing:

- Meletus looks like he is "making a test" (*diapειρômenô(i)*) in his indictment of Socrates "to see if Socrates, the 'wise man,' will recognize the contradiction in his jest" (*Ap.* 27a; see also *apopeirômenos*, 27e).
- Prodicus is testing (*apopeirasthai*) Protagoras (*Prt.* 341d) when Prodicus falsely says that Simonides meant "bad," not "difficult," by "hard," "in order to see if Protagoras can defend his statement."
- Protagoras was likely to have been testing (*apopeirômenos*) Socrates when Protagoras [falsely] said that virtue had parts (*Prt.* 349c).

There is good reason for Socrates to use this false-lead test in his pedagogy. In the first place, although he himself does not believe that any but God know what virtue is (*Ap.* 23a), he can determine if his interlocutors take themselves to be knowledgeable by falsely claiming or suggesting that his interlocutors have such knowledge. If his interlocutors pretend to knowledge, an additional false lead accepted by them will make it possible for Socrates to drive them into contradiction. I can think of no better way than this procedure, repeated as often as necessary, to help those of us at the lowest level recognize our own ignorance. And it is sound pedagogy for Socrates to refrain from eliciting positive doctrines from his interlocutors until they are at the middle level, by a principle of priority: interlocutors at the lowest level of attainment

of wisdom need to reach the middle level much more than they need a few more premises and conclusions conjoined to their conceit of wisdom!<sup>14</sup>

In what follows, I show that Socrates recycles four pedagogical steps with interlocutors at the lowest level of attainment of wisdom:

1. (*Question*) Socrates asks the question.
2. (*False lead*) If necessary, he prompts the interlocutor with a suggestion that is false, but seems true to those who are ignorant.
3. (*Accepted*) The interlocutor (unless he needed no prompting) accepts Socrates' suggestion.
4. (*Refutation*) Socrates attempts to refute the interlocutor's answer.

If Socrates succeeds in refuting the interlocutor, he repeats his original question and the process is repeated, as often as necessary, until the interlocutor either breaks off the discussion or admits his ignorance.

## 10.6 Slave-Boy Illustration

I begin with the slave-boy discussion in the *Meno* to illustrate Socrates' use of these four steps. I do not, in using this passage, make any claims about early vs. middle dialogues or about whether the Socrates who argues for RMI also argues for a theory of recollection. My thesis only requires that the slave-boy discussion, at least until the boy reaches the middle level of attainment of wisdom, is an apt illustration of how Socrates may test his interlocutors in the RMI dialogues. The subject, geometry, is something both Socrates and his readers know, so that it makes a clear illustration for this sort of Socratic test: no one will suggest that Socrates himself favors one of the false leads on the grounds of its seeming truth. The boy, Socrates points out, begins at the lowest level of attainment of knowledge about the square, "thinking he knows when he does not know" (82e). We find the expected steps of the testing process:

- (*Question*) Given a square of area 4, "Try to tell me how long each side of the square of area 8 will be" (82de).
- (*False lead*) "The side of *this* square is 2 feet. What will be the side of the twice-as-big square?" (82e). Socrates' suggestion to look to the *side* of the square of area 4, rather than its *diagonal*, is misleading, as is his suggestion that the length of that square's side, 2, is relevant to the twice-as-big-area problem.<sup>15</sup>
- (*Accepted*) "Obviously, Socrates, twice as big" (82e).
- (*Refutation*) Socrates questions the boy step-by-step (*ephexês*) until the boy retracts his answer (82e–83c).
- (*Question*) Socrates repeats the original question (82c).
- (*False lead*) Socrates points out that doubling the side of the 4-foot square produces a 16-foot square, that the desired 8-foot square is double the 4-foot square

but half the 16-foot square, and that the 8-foot square will have a side bigger than 2, but smaller than 4 (83c–d). Socrates misleads by continuing to direct the boy’s attention to the *sides* of the squares of area 4 and of 16, and by suggesting that some simple relationship between those sides is relevant to the solution.

- (*Accepted*) The boy seems influenced by Socrates’ suggestion, saying that the
- 8-foot square will have a side of 3 feet (83e).
- (*Refutation*) Socrates again cross-examines until the boy retracts his answer
- (83e).
- (*Question*) Socrates again repeats his original question (83e).

At this point, the boy describes himself as ignorant in strong terms: “By God, Socrates, *I* do not know” (*alla ma ton Dia, ô Sôkrates, egôge ouk oida*, 84a). He has reached the middle level, knowing his own ignorance and “happy to look” (*zêtêseien an hêdeôs*, 84b) for knowledge (84a–c).<sup>16</sup> Socrates thereupon questions the boy further, demonstrating that the square of area 4 has a side equal to the diagonal of the square of area 2 (84d–85b), leaving the boy in a “dreamlike” state (86d). Socrates says that “if the same questions were put to him on many occasions and in different ways [by someone who knows geometry], in the end he would have as much knowledge as anyone on the subject” (85c–d).<sup>17</sup>

We should not expect this eliciting of positive doctrine to occur with Laches, Nicias, Meno, or Euthyphro, since none of these ever reach the middle level. The subject in the *Laches*, *Meno*, and *Euthyphro* is virtue, which, unfortunately, is not as well known as geometry, so that I cannot assume a consensus among us interpreters on the subject. However, if we accept the hypothesis that RMI is true, at least by Socrates’ lights, we find parallels between these three dialogues and the slave-boy passage of the *Meno*.

## 10.7 Application to *Laches*, *Meno*, and *Euthyphro*

Consider the *Laches*, where Socrates tests both Laches and Nicias. The *Laches* begins with Nicias (at 180a), Laches (at 180b), and Socrates (181d) consenting to help some old friends by advising them about the proper way to raise their children. Unfortunately, Nicias and Laches disagree in their advice, and Socrates is called upon to cast the deciding vote. Instead of voting, Socrates begins to cross-examine the party. Nicias, who knows Socrates best, tells the party of his familiarity with Socrates’ techniques and predicts that Socrates “will not let them go until he tests (*basanisê(i)*) them well” (188a). All agree that the question is “how virtue may come to the souls of their children and make them better” (*tin an tropon tois huesin autôn aretê paragenomenê tais psuchais ameinous poiêseie*, 190b). Thus Socrates elicits agreement that “we need to know what virtue is” (*oun hêmîn touto g’ huparchein dei, to eidenai hoti pot’ estin aretê*, 190b)—“because, I suppose, if we did not thoroughly know what virtue actually is” (*ei gar pou mêd’ aretên eideimen to parapan hoti pote tugchanei hon*, 190b) we could not be advisers about the best way to



acquire it. After Laches agrees to this, Socrates makes the assertion, “Therefore we claim to know what virtue is” (*phamen ara. . . eidenai auto hoti estin*, 190c).

My interpretation, according to which Socrates for pedagogical reasons sometimes makes false leads to test his interlocutors’ pretenses to wisdom, can explain his claim to know at 190c: he is testing Laches and Nicias, as Nicias predicted and as the illustrative dialogue with the slave boy in the *Meno* would lead us to expect, with just the sort of test that Socrates has described others as making.<sup>18</sup> The only alternative I can see to my false-lead interpretation is to give up on the project of a coherent account of Socratic moral theory, not only between the *Laches* and the *Apology*, but within the *Laches* itself, for even in that dialogue Socrates makes his characteristic denial of knowledge of virtue, already at 186c and also at the conclusion of the dialogue.

We find the same steps of the testing process, with Laches beginning at the lowest level: “Yes [we know what virtue is]” (190c); “It is easy to say what courage is” (190e).

- (*False lead*) “I would not have us begin with the whole, . . . let us begin with a part [of virtue, courage]” (190d) = PW/L.<sup>19</sup>
- (*Accepted*) Laches says, “Yes, certainly” (190d).
- (*Question*) What is courage? (190e).
- Laches needs no prompting from Socrates, and begins by giving a specific instance instead of a general definition of courage.
- (*Refutation*) Socrates gets Laches to see that Laches has not answered the “What is courage?” question by saying “Courage is staying at one’s post” (190e–191e).
- (*Question*) Again, “what is that common quality which is called courage?” (192b).
- Laches again takes no prompting: “Courage is endurance of soul” (192b).
- (*Refutation*) Socrates gets Laches to revise his definition to “Courage is wise endurance of soul” (192d), which he goes on to refute (192e–193e).

At this point in the dialogue, Laches has contradicted himself, but he blames the *aporia* in the discussion upon something other than his own ignorance: “I am unused to this sort of inquiry. . . I am really grieved at being unable to express my meaning, for I think I do know (*noein men gar emoige dokō*) the nature of courage” (194a–b). If we expect Socrates to be engaged in the pedagogy of moving Laches and Nicias from the lowest to the middle level of attainment of goodness and wisdom, we may interpret his claim (PW/L) that courage is but a part of virtue to be a test for them and to cast no doubt on the hypothesis that Socrates’ moral theory is RMI.

Socrates next turns to Nicias, and we find the same steps of the testing process:

- (*Question*) “Tell us, Nicias, what you think about courage” (194c).
- (*False lead*) Laches proceeds to question Nicias until Socrates takes over, and Socrates begins with a false lead: “You yourself said that courage was a part, and there were many other parts, all of which taken together are called virtue. . . . In that case, do you say the same as I? In addition to courage, I call temperance,

righteousness, and the like virtue. Would you not say the same?" (198a–b) = PW/N.<sup>20</sup>

- (*Accepted*) "Certainly" (198a, b).
- (*Refutation*) 198a–199e. (See Section 10.2 above for discussion.)

Socrates does not get a chance to repeat his question to Nicias, for Laches interrupts with abusive remarks. Nicias in reply quits the testing process: "Enough has been said on the subject" (200b). Again, if we expect Socrates to be engaged in the pedagogy of moving Laches from the lowest to the middle level of attainment of goodness and wisdom, his claim (PW/N) that courage is but a part of virtue casts no doubt on the hypothesis that Socrates' moral theory is RMI.

Consider next Meno, who also begins at the lowest level of attainment of wisdom, taking himself to know what virtue is: "It is easy to say what virtue is" (71e). Again we find the false-lead testing process.

- (*Question*) "What is virtue?" (71d).
- Meno needs no prompting from Socrates, and begins by giving a specific instance instead of a general definition of virtue (71e–77b).
- (*False lead*) When Meno says "justice is virtue," Socrates asks him, "Is it virtue or a virtue?" Then Socrates gives a roundness/shape analogy, suggesting that the justice/virtue relation is like roundness/shape (73e, repeated at 78d) = PW/M.
- (*Accepted*) Meno in saying justice is one among many other virtues takes himself to be agreeing with Socrates (73e).
- (*Refutation*) Socrates gets Meno to see that Meno has not answered the "What is virtue?" question by saying virtue is justice, temperance, etc. (74a–77a).
- (*Question*) Socrates repeats his question (77a).
- Meno needs no prompting; he says virtue is "desiring fine things and being able to get them" (77b).
- (*Refutation*) Socrates shows all are alike in desiring the good, so revises definition to "power to get good things" (77b–78b).
- (*False lead*) Socrates suggests that good things include health and wealth (78c).<sup>21</sup> I call this a false lead because Socrates is well aware that health and wealth are not goods, since without use they are no good at all and misused are bad (*Euthd.* 280d–281e).
- (*Accepted*) Meno agrees and adds, "high and prestigious office in the state" (78c).
- (*Refutation*) Socrates cross-examines until Meno retracts his answer (78d–79d).
- (*Question*) Socrates repeats his original question again (79e).

Meno, unable to answer, blames the *aporia* in the discussion upon something other than his own ignorance: "Socrates, you are exercising magic and witchcraft upon me and positively laying me under your spell until I am a mass of helplessness" (80a).<sup>22</sup> He has not reached the middle level, nor should we expect Socrates to have been developing positive doctrines with his claims either (PW/M) that virtue contains parts or that good things include health and wealth. Notice that Meno, like Laches, does not get as close to RMI as Nicias, who was a reductive intellectualist.

Accordingly Socrates, although ready in case Meno reached intellectualism in the discussion, does not get to use his false suggestion that virtue has parts. I suppose that had Meno (or Laches) managed eventually to define virtue as wisdom, Socrates would have used the suggestion that virtue has parts to refute him.

Consider, finally, Euthyphro, who also begins at the lowest level, thinking himself wise about the nature of piety (4e–5a, 5c). Again we find the predictable pedagogical pattern:

- (*Question*) “What is piety?” (4c).
- Euthyphro needs no prompting but gives specific instances instead of a general definition of piety (4d–6e).
- (*Refutation*) Socrates gets Euthyphro to see that he has not answered the ‘What is piety?’ question by giving instances of piety (6d–e).
- (*Question*) “Teach me what the form [of piety] is!” (6e).
- Euthyphro again needs no prompting: “Piety is what is pleasing to the gods” (6e–7a).
- (*Refutation*) Socrates gets Euthyphro to revise his definition to “Piety is what is pleasing to all the gods” (7b–8a), then gets him to agree that even the revised definition fails (9c–11a).
- (*Question*) “Begin again: say what the holy is” (11b).

At this point, Euthyphro is unable to answer. But he does not reach the middle level of coming to see his own ignorance. Instead, he blames Socrates for his own failure to answer: “Socrates, I’m not able to tell you what I have in mind. . . *you* are the Daedalus [making the assertions move out of place]” (11b–d).<sup>23</sup> Accordingly, our model of Socratic pedagogy predicts Socrates will not start to derive positive theorems about piety, but rather will keep giving him false leads until Euthyphro reaches the middle level. This is exactly what we find, if we assume the RMI interpretation to be correct:

- (*False lead*) When Euthyphro agrees that all that is holy is just, Socrates asks him, “And is all justice holy, too?” Then Socrates gives the fear/shame analogy, suggesting that justice is to piety as fear is to shame (12d) = PW/E.
- (*Accepted*) Euthyphro takes himself to be agreeing with Socrates’ own view: “That is my opinion; I think that you, Socrates, are clearly right” (12d).
- (*Refutation*) Socrates attempts but does not complete a refutation (12e–14b).

Socrates is frustrated by Euthyphro’s answer to his question, “What is the greatest result that the gods produce when they employ human beings in their service?” (as his remarks at 14b–c show). Euthyphro has a tendency to assimilate piety with justice: his star instance of piety is “opposing [= “prosecuting”] unjust people (*tô(i) adikounti. . . epexienai*, 5d). And Euthyphro’s assimilation must have been natural for Greeks, for Socrates relied upon it in his argument identifying piety and justice at *Prt.* 330a–331b. Given this tendency, Socrates might well have expected Euthyphro to have answered the question how we serve the gods by saying, “We serve the gods

not by acts of shipbuilding, winning victory in war, or growing food, but by acts of justice.” This answer would have enabled Socrates to go on to reach a contradiction with his earlier false suggestion that piety is but a part of justice. Euthyphro, frustratingly, does not get the point. Socrates, indefatigable, tries again:

- (*Question*) “Once more, how do you define the holy?” (14c).
- Euthyphro needed no prompting to suggest a new answer before the question was even asked, “[Piety is] knowing. . . how to pray and sacrifice” (14b).
- (*False lead*) “And to give properly to the gods is to give them those things which they happen to need to receive from us?” (14e: I call this suggestive question a false lead because they agreed just earlier, 13c, that the gods can gain no benefits from us).
- (*Accepted*) “True, Socrates” (14e).
- (*Refutation*) Piety becomes a business transaction, in which humans provide advantage to the gods, which Euthyphro recognizes is untenable: “What!—Socrates, do you suppose that the gods gain anything by what they get from us?”

Socrates can now raise again the question—“What, then, are the gifts we give to the gods?”—to which we might expect a Greek such as Euthyphro to answer, “Our proper gift to the gods consists in our human acts of justice to each other.” This answer would allow Socrates to use his earlier false lead, that piety is but a part of justice, to try to refute Euthyphro and bring him to the middle level. Euthyphro does not find this answer, and leads himself into a different refutation (15b–c). And so Socrates begins another cycle of the testing process with the question: “So we must go back again, and start from the beginning to find out what the holy is” (15c). Euthyphro here breaks off the conversation and hurries off (15d–e), never admitting his ignorance or reaching the middle level, despite Socrates’ best efforts. Once again, Socrates’ suggestion (PW/E) that piety is part of justice casts no doubt on the RMI interpretation, just as his suggestion that the gods need our gifts need not conflict with other passages, in the *Euthyphro* itself, where Socrates indicates that the gods are perfect (6a–c, 13c).

## 10.8 Conclusion

On my account, there are dialogues that show us a Socrates engaged in the pedagogy of helping move others from the lowest to the middle level of goodness or wisdom. Those at the lowest level seem to know, though they are ignorant. Socrates describes himself as testing such people. He is aware of the sort of test where one suggests a false lead to see if the other will recognize its falsity, and we see him using just such false leads in his cross-examination of the slave boy. We find false leads—that Socrates, like most people, knows what virtue is, that health and wealth are good things, that the gods need things—in his discussions with Laches, Nicias, Meno,

and Euthyphro, whether or not we accept the RMI interpretation. If we suppose that Socrates held the RMI view of virtue, then among the false leads is the suggestion or claim that piety is but a part of virtue. Those claims and suggestions (namely, PW/L, PW/N, PW/M, PW/E) cast no doubt upon the interpretation that Socrates was a reductive monist intellectualist about virtue. I have therefore disarmed any objections to the RMI interpretation based upon those passages.

Let me consider one final objection. My argument, that Socrates makes claims and suggestions that are by his own lights false, is self-defeating as a defense of the RMI interpretation. For the only reasonable conclusion, if my pedagogical account is correct, is complete skepticism about the views of the Socrates of these dialogues, since no aporetic dialogue can contain any statement that counts as evidence for Socrates' views.

In reply, I readily admit that no statement taken in isolation from these dialogues can serve as evidence. What saves me from exegetical skepticism are not isolated statements but lines of argument, as sketched in Section 10.2, that lead to RMI. By contrast, *in none of these dialogues is there even one argument driving us towards a part/whole account of virtue*. Our task, then, to avoid skepticism about Socrates' moral theory, is to examine his arguments to see if they are as laughable as they at first seem or if, when closely examined, they prove to be compelling. The consequence of this approach is that the knowledge how to interpret Socratic texts is the very same as the knowledge of the truth about such things as piety.<sup>24</sup>

## Notes

1. Santas (1964) defended the philosophical viability of Socratic intellectualism. His paper was the seed from which has grown a flower of philosophically astute scholarship on Socratic intellectualism. The flower has two branches: Vlastos (1972) interpreted Socrates as a reductive *pluralist* intellectualist about virtue; Penner (1973) a reductive *monist* intellectualist, starting the scholarly conversation in letters to which this paper contributes.
2. In *Religion Within the Limits of Unassisted Reason*, Kant writes:

Religion is the recognition of all our duties as divine commands. . . . In a universal religion there are no special duties towards God. . . . If anyone finds such a duty in the reverence due to God, he does not reflect that this is no particular act of religion, but a religious temper accompanying all our acts of duty without distraction.

- Among Hebrew prophets, Isaiah 1:11–17 says God wants justice, not ritual offerings, likewise Micah 6:6–8 and Psalm 50 (citations from Irwin 1989, p. 77 n. 16). Jesus summarizes the Ten Commandments as but two: love God and love others (Luke 10:27), and in the Parable of the Last Judgment makes clear that we love God by loving others (Matthew 25:32–46), which allows him to reduce divine commandments to only one Great Command, to love others (John 15:12, 17).
3. Calef (1995, p. 9 n. 30), defending the RMI interpretation, grants that Socrates in his cross-examination at PW/E may allow Euthyphro to “flounder” but denies that Socrates misleads. He does not discuss PW/L, PW/N, or PW/M. McPherran (1985, p. 286), critical of the RMI interpretation, thinks that if Socrates misleads, he is blameworthy for sophism and “trickery.” He does not consider the possibility of pedagogical value in false leads.

4. Rickless (1998, p. 359) gives the following additional objection to the RMI interpretation: RMI makes Socrates' position in the *Protagoras* incoherent, because Socrates argues that

(1) courage is identical to wisdom (i.e., knowledge)  
and that

(2) courage is identical to knowledge of what is to be feared and dared.

I reply that (1) and (2) are incoherent only if we restate (1) as

(1') courage is identical to wisdom per se (that is, to the genus wisdom or perhaps the sum of all kinds of wisdom and *not* one species of wisdom only),

but there is no reason why RMI cannot restate (i) in the consistent form of

(1'') courage is identical to wisdom of a kind.

This restatement of (1) as (1''), which Rickless himself accepts (*ibid.*, p. 362), disarms his objection.

Woodruff (1976) attempts to reconcile the seeming contradiction between the statements that courage is only a part but also all of virtue. Woodruff distinguishes courage in essence (*courage-itself*) from courage in accident (*courage-in-ingots*, as it were, on the model of the distinction between *the substance gold*, which is all one, and *gold-in-ingots*, which has as many parts as there are ingots). According to this distinction, some things can be true of courage-in-ingots that are not true of courage-itself. For example, it may be that the predicate *knowledge only of future goods and evils* is true of courage-in-ingots but false of courage-itself.

But, as it seems to me, Woodruff's distinction cannot escape Socrates' argument. To see why, let us accept Woodruff's distinction and make Woodruff's assumption that when Socrates speaks of courage as a part of virtue, he is speaking of courage-in-ingots, not courage itself. Accordingly, courage-in-ingots is a part of virtue. Moreover, as Socrates and Nicias agree, this very courage-in-ingots will be nothing but the knowledge or science of future goods and evils. And it is undeniable, as Socrates and Laches rightly agree, that *there is no distinction* between the science that knows future goods and evils and the science that knows goods and evils past, present, and future (*La.* 198d, 199a). Thus in whatever sense courage-in-ingots is the science of future goods and evils, it is precisely also the knowledge of all goods and evils, past, present, and future. It follows that courage-in-ingots is the whole of virtue, which contradicts Woodruff's interpretation.

5. Rudebusch (1999, pp. 108–13) defends the inference from virtue being knowledge of living well to a good man being invulnerable.
6. I follow Santas (1969, pp. 197–202), who defends Socrates' argument against the pluralist doctrine that courage is but a part of virtue. Rickless (1998, pp. 361–2) and McPherran (2000, p. 313) interpret Socrates as a pluralist holding versions of RI, CP, and CF. Since Socrates rightly uses UK to *refute* just such pluralism, I take it that his argument refutes as well these two pluralist interpretations.
7. Rickless (1998, p. 362), following Ferejohn (1984, p. 384), tries to distinguish temperance from justice on the grounds that temperance is the knowledge of what is good and bad *for oneself* while justice is the knowledge of what is good and bad *for others*. But the very same knowledge knows both—for example, it would be absurd to claim, Ion-like, to know how to heal not all human bodies, but only one particular human body! Again, the same authors, on the same pages, try to distinguish piety as a mere part of justice on the grounds that piety is knowledge of what is good and bad *for the gods*. But they must face a dilemma: if the gods are specifically the same as human beings with respect to their good and bad, then piety is precisely justice, while if the gods specifically differ from human beings, then piety will no more be a part of justice than the knowledge of horse training. Indeed horse training will be more like justice than piety, since horses, like human beings, have imperfections to remedy; the gods have no such imperfections and consequently their good and bad must be generically different from horse and human good and bad.
8. It is a mistake, therefore, to interpret Socrates in the *Protagoras* to hold that courage, piety, etc. are mere parts of virtue in the way a lump of gold has parts, as do e.g., Vlastos (1972, p. 230), McPherran (2000, p. 313), and Woodruff (1976, p. 102). Socrates is undeniably surprised at

any account that makes “justice and temperance and piety and all these things, *taken together*, one thing, virtue”—whether the account is in terms of parts of a face or parts of a lump of gold makes no difference to him. Protagoras chooses the parts-of-face version, and accordingly Socrates’ arguments are directed against that version. Had Protagoras chosen the parts-of-lump analogy, Socrates could have used arguments from the identity conditions of knowledge to refute him, as shown above, note 7.

9. The *Theaetetus* mentions additional advantages of the middle level in comparison to the lowest level. Thanks to Socrates’ “examination” (*exetasin*, 210c), any future investigations we make will be “better” (*beltionôn*, 210c), and even if we never escape our ignorance, our knowledge of it will make us “more civilized and easier for our companions to bear” (*hêttôn . . . barus tois sunousi kai hêmerôteros*, 210c).
10. The point that in this passage Socrates is comically outdoing the sophist at sophistry may be true; it does not follow that the doctrines Socrates attributes there to Simonides are not Socrates’ own.
11. At *Euthydemus* 281b–e Socrates explains why ignorance at the skill of human goodness will negate the advantages of other skills: a fool will get no benefit from doing many things; on the contrary, if a fool does less he will make fewer mistakes and be less wretched.  
Brickhouse and Smith (1994, p. 17) mention an additional disadvantage to the lowest level: Socrates’ method of cross-examining others brings to light contradictions in the beliefs of those at the lowest level. “If one’s beliefs about how it is best for one to live are inconsistent, one cannot . . . follow all of one’s . . . inclinations; in such a condition, one will be doomed, at least to some degree, to a life of frustration and inner conflict.”
12. See Brickhouse and Smith (1994) and Benson (2000) for accounts of Socratic cross-examination.
13. I agree with the point, made for example by Brickhouse and Smith (1994, p. 15), that “Socrates does not always accept the truth of the premises he uses.” They prove the point by referring (1) to instances where Socrates uses *reductio ad absurdum* (“indirect proof”) and (2) to Socrates’ use in the *Euthyphro* (6b–8b) of Euthyphro’s belief, which Socrates does not share, that the gods quarrel and disagree. I go beyond this point in my account of pedagogical testing.
14. Hugh Benson in correspondence has described my account of false-lead pedagogy as a version of his non-constructive account of the elenchus (see Benson 2000, pp. 35–36), adding to non-constructivism the “presumption that we can tell ahead of time—presumably in light of a comprehensive interpretation of the dialogues—which of the beliefs of the interlocutors Socrates thinks are false.” Benson’s non-constructivist thesis (p. 35) is that “Socrates neither can nor does conclude as a result of an individual elenctic episode the falsehood of the apparent refutand.” As Benson notes (p. 35 n. 14), “the only [account] that is directly at odds with [his] is the account according to which Socrates concludes *as a result of an individual* elenctic episode that the apparent refutand is false and he is justified in doing so”—he is “no longer sure that anyone ever wanted to recommend such an account.”

How can we tell which beliefs Socrates will think are false? In answer to this question, I have two points to make. First, according to the RMI interpretation (evidence for which is surveyed in Section 10.2 of this chapter), Socrates will think any non-monist account or non-intellectual account of virtue is false (eliminative accounts of the virtues disagree with Socrates’ reductive account only in terminology). Socrates would of course also recognize that the majority of people, then as now, are inclined to assent to non-monist and non-intellectual accounts of virtue. Second, according to the three-levels-of-wisdom pedagogical assumption I attribute to Socrates on the basis of Sections 10.3, 10.4, and 10.5 of this chapter—which I intend to be far less controversial than the RMI interpretation!—Socrates will be quick to test people to see if they assent to statements that they possess knowledge of human virtue.

My account here is also in agreement with Brickhouse and Smith (1994), who argue that Socratic conversation has a constructive function in two senses: first, “Socrates’ contributions to the conversation do betray his own beliefs” (p. 65), second, “to the extent that he has generated inductive evidence through previous elenctic examinations for the necessity of his own



view for a coherent life, Socrates can claim to have established a general truth applicable to all” (p. 19). As Benson and Brickhouse and Smith all recognize, their non-constructivist and constructivist theses are mutually compatible (Brickhouse and Smith 1994, p. 12 n. 18; Benson 2000, p. 35 n. 14).

15. As I recollect, Terry Penner pointed out the false leads Socrates suggests to the slave boy in a seminar in the late seventies. Penner’s point was that Socrates’ leading questions discourage the boy from acting as a sycophant, as one might expect from an enslaved person. For, if it is the case that the boy’s answers at 82e2–3 (“Twice the length”) and 83e2 (“Three feet”) are attempts at sycophancy, those answers fail to flatter Socrates, since the boy must retract both answers.

I endorse Penner’s point about Socrates asking questions that discourage sycophancy. My point is that there is an additional purpose to Socrates’ suggestions. His suggestion will only be taken to a false conclusion by a person who is both ignorant and is ignorant of his own ignorance. Precisely for this reason his false leads do not make Socrates a liar, that is, one who deliberately tries to mislead, but rather an expert at testing for ignorance in others by asking leading questions. (My thanks to Betty Belfiore for discussion on this point.)

16. “Before conversing with Socrates, the interlocutor may have felt no particular need to mend his ways. But the *aporia* that results from Socratic questioning gives the interlocutor an important reason to pursue the examined life: the recognition that one is seriously confused about how best to live” (Brickhouse and Smith 1994, p. 17).
17. I endorse Benson (2000, p. 259), who distinguishes four stages in Socrates’ conversation with the slave boy. First, “Socrates asks and explains the question.” Second, “Socrates refutes the slave boy’s initial answers. . . [until] the slave boy responds that he does not know.” Third, “Socrates leads the slave boy. . . to. . . belief.” Fourth (this stage is described but not illustrated by Socrates), “the slave boy would be led from true belief to knowledge.” And Benson on the same page makes the point that is essential to my thesis: “it is not until the third stage of the slave-boy conversation that the substantive theory” is elicited.
18. Again (see n. 15 above) I do not claim that Socrates tells a lie when he tests Laches in this manner. Socrates points out to Laches that implicit in the act of advising is a claim to know (190b7–c2). Laches agrees that there is such an implication in the act of advising *and does not express any reservations about the advising he has done or proposes to do* (190c3). From this fact Socrates infers (“therefore”) what must follow from Laches’ agreement in dramatic context: “[Since we are going to be advisors,] therefore we say that we know what virtue is.—Of course we say so” (190c4–5). Socrates’ statement of group knowledge (“we”) is not a lie about his own state of conscious ignorance but the first premise in what will be a *reductio ad absurdum*. (Again I thank Betty Belfiore and Gail Fine for making this clear to me.)

My false-lead hypothesis explains another statement of Socrates, where he suggests that Nicias and Laches must be wise:

If Nicias or Laches has discovered or learned it, I would not be surprised; for they are wealthier than I am, and may therefore have learned of others, and they are older too, so that they have had more time to make the discovery. They do seem able to educate a man, for unless they had been confident in their own knowledge, they would never have spoken thus unhesitatingly of the pursuits which are advantageous or hurtful to a young man. (186c5–d3)

It is uncharitable to take this speech to be mean-spirited sarcasm. On the contrary, it is evidence of aristocratic good manners, which are governed by the general principle of charity, in this particular case requiring Socrates to assume the best of others and hence to interpret their knowledge claims in the best possible light. Moreover, although his suggesting there is a likelihood of wisdom in his companions is well-mannered, the suggestion also serves as a test, giving his companions an opportunity to accept or reject the suggestion of their wisdom.

19. Again Socrates is not telling a lie here. He hedges his words: to inquire into the whole of virtue is “perhaps more work” (*pleon. . . isōs ergon*, 190c9) than to inquire into a part. It is

- more accurate to describe him here as opening a door, so to speak, to see if his conversation partner will walk through it: a test.
20. Again Socrates does not tell a lie, since he himself does call courage, temperance, righteousness and the like each virtue. It is more accurate to describe him as putting before his conversation partner a test.
  21. Again Socrates does not lie but tests: "And don't you call such things as health and wealth goods?" (78c5–6).
  22. Meno, in not recognizing his own ignorance, does not reach the slave boy's level of self-understanding. This difference between Meno and the slave boy is unnoticed by Benson (2000, p. 259), who states that "Socrates understands these first two stages of the conversation to be an exact parallel to the main conversation with Meno. . . . By the end of the [stage where Socrates has refuted his initial answers] the slave boy has reached the precise point Meno had reached."
  23. Brickhouse and Smith (1994, p. 22) seem to confuse the arrival of Euthyphro at *aporia* with his attainment of the middle level. For they take the fact that "Euthyphro has been fully reduced to *aporia* by *Euthyphro* 11b6" to entail that Socrates has no further work to do to bring Euthyphro to the middle level, and hence that any subsequent examination of definitions is part of the "positive" process of moving from the middle level to the highest level, a process which we ought to expect to be free of false leads: "[After 11b6] Socrates continues his search for the definition of piety by contributing one of what certainly appears to be his own views. . . that piety is a part of justice."
  24. As it seems to me, Santas's work is a paradigm of this identity of interpretation and philosophy. I thank participants in the 2001 Arizona Plato Colloquium, especially my commentator Joel Martinez, for helpful discussion and comments, and also Betty Belfiore for correcting many errors in correspondence in 2005.

## Bibliography

- Benson, H., ed. 1992. *Essays on the Philosophy of Socrates*. Oxford and New York: Oxford University Press.
- Benson, H. 2000. *Socratic Wisdom*. Oxford and New York: Oxford University Press.
- Brickhouse, T., and N. Smith. 1994. *Plato's Socrates*. Oxford and New York: Oxford University Press.
- Calef, S. 1995. "Piety and the Unity of Virtue in *Euthyphro* 11e–14c." *Oxford Studies in Ancient Philosophy* 13:1–26.
- Ferejohn, M. 1984. "Socratic Virtue as the Parts of Itself." *Philosophy and Phenomenological Research* 44:377–88.
- Irwin, T. 1989. *Classical Thought*. Oxford and New York: Oxford University Press.
- McPherran, M. 1985. "Socratic Piety: In Response to Scott Calef." *Oxford Studies in Ancient Philosophy* 13:27–36. Reprinted in Benson (1992).
- McPherran, M. 2000. "Piety, Justice, and the Unity of Virtue." *Journal of the History of Philosophy* 38:299–328.
- Penner, T. 1973. "The Unity of Virtue." *Philosophical Review* 82:35–68. Reprinted in Benson (1992).
- Rickless, S. 1998. "Socrates' Moral Intellectualism." *Pacific Philosophical Quarterly* 79:355–67.
- Rudebusch, G. 1999. *Socrates, Pleasure, and Value*. Oxford and New York: Oxford University Press.
- Santas, G. 1964. "The Socratic Paradoxes." *Philosophical Review* 73:147–64. References are to revised version in Santas (1979).
- Santas, G. 1969. "Socrates at Work on Virtue and Knowledge in Plato's *Laches*." *Review of Metaphysics* 22:433–60. References are to the reprint in Vlastos (1971).
- Santas, G. 1979. *Socrates: Philosophy in Plato's Early Dialogues*. London: Routledge & Kegan Paul.
- Vlastos, G. 1971. *The Philosophy of Socrates: A Collection of Critical Essays*. New York: Anchor.

- Vlastos, G. 1972. "The Unity of the Virtues in the *Protagoras*." *Review of Metaphysics* 25: 415–458. Reprinted in Vlastos (1981).
- Vlastos, G. 1981. *Platonic Studies* (2nd ed.). Princeton: Princeton University Press.
- Woodruff, P. 1976. "Socrates on the Parts of Virtue." *Canadian Journal of Philosophy* 2(Supplementary):101–116.

# Chapter 11

## The *Republic* as Philosophical Drama

John P. Anton

### 11.1 The Antecedents

The aftereffects of the Peloponnesian War were still spawning the pains and failures of the Greek cities making even more obvious the need for political decisions to prevent another yet destructive conflict. Plato was still in his early manhood when it became clear to him that the continuation of political decadence called for careful diagnosis if it was ever to be stopped or at least subjected to reasonable controls. One of Plato's motives for turning to philosophy as well as writing the *Republic* was the search for such political wisdom, so that at least the execution of philosophers would not be repeated.

By writing the *Phaedo*, Plato made clear what conceivably transpired in the prison on the last day of Socrates' life. At the same time, however, he was issuing a statement on his own principles concerning the best and just type of life available to all citizens and all human beings in general. The *Phaedo* defined a commitment as well as a credo to the highest attainments of the reflective life. As such these principles were defended as the foundations for a life truly worth living. In this regard, by writing the *Phaedo* as the last will and testimony of Socrates, Plato came to declare what he could not ignore in his later compositions or repudiate and perchance compromise. If such be the case, it follows that the *Republic* has to be read as a continuation as well as confirmation of the doctrines so vividly stated in the *Phaedo*. The drama of the *Phaedo* is preserved in the *Republic*, while the latter, although written after the former, remains a retrospective work anticipating what will be the epilogue of Socrates' career. As dramatic setting, the *Republic* shows Socrates at the age of sixty exploring the grounds for securing the meaning of justice, individual and political, while the War was still going on and the destruction was getting closer to its final phase. In the midst of all the turmoil, Socrates and his companions will once again return to the philosophical quest of justice. Needless to

---

J.P. Anton (✉)  
University of South Florida, Tampa, FL, USA  
e-mail: hanton1@tampabay.rr.com

say, decades later, when Plato will write the *Republic*, the quest was still timely as well as needed to illumine the human condition.

When Plato taught that living is the function of the soul, he was establishing as well as clarifying a long tradition that had sought to throw light on the nature of human existence. As the discussion in the opening of the *Republic* progressed and the search for the definition of justice reached a critical point, the nature of the soul and its functions, including those of its parts, became pivotal to the next phase. But to talk about the functioning of anything, be it a part of the body, say the eye, is far from giving a complete account without reference to the proper functioning organ at its best. Thus it is not enough to say that the functioning of the eye is seeing and that of the soul is living unless it is shown how the eye sees well and the soul lives best. Briefly put, a part, be it of the body or of the soul, performs its function well when it can be shown to possess its own excellence while it is at work (*oikeia aretê*). If not, the eye or the soul, including the parts of the latter, function defectively from its own badness (*oikeia kakia*). The same pertains to the body in its entirety and to the soul as a whole. The normal course, then, is to expect the soul to function properly by performing excellently and in accord with its own *aretê*. Whereas identifying the excellent functioning of the eye presents no difficulty, determining that of the soul amounts to a formidable undertaking. At *Rep.* 353d Socrates focuses on the complex question of the soul's excellence and comes to assert that it cannot be properly identified without introducing the concept of justice with reference to the soul in its entirety and its parts. What is needed is a clearly stated theory of justice to conclude the refutation Socrates and his friends had started of Thrasymachus' position. At this point Adeimantus states his expectations:

Since you've agreed that justice is among the greatest goods, which are worth possessing for their consequences but much more for themselves alone—like seeing, hearing, thinking and being healthy, and all other goods genuine in their own nature rather than by appearance—please then praise this very thing about justice: how justice in and of itself benefits him who has it, and injustice harms. (367c–d; Allen transl.)

Socrates accepts the assignment but instead of coming forth with a theory to meet Adeimantus' challenge directly, he postpones the answer and asks him to join him in constructing a just *polis*, not a just individual, whereby justice will be writ large and become identifiable. It is understood that they would be constructing a model of justice on the basis of an analogy. They could then use the construct to see what the just soul of an individual citizen is. The move is well conceived. The discussion proceeds to explore first the elements needed for a theory of human nature. The results can then be used as the basis for understanding the complexities of political life. However, the search that had just begun will soon reveal that the parts of the soul may and frequently do function at cross purposes and in sharp conflict with each other thus pulling the individual in opposite directions at the same time. More often than not this diversity operates at the expense of the soul's main function, living well and in accordance with its own proper excellence. The consequences of the diversity of functions of the parts of the soul will become sharply visible at another level: the political life. Here is where the problems multiply and the difficulties of

maintaining community cooperation tend to cause the evils of sharp divisions and even bloodshed. Defining justice, therefore, is better done not with the selection of materials from actual situations in history but with the guidance of a theoretical model once the understanding of human nature is in place. Adeimantus and Glaucon promise their cooperation.

Unless the soul is efficient in all of its functions, expecting the *polis* to perform well is impossible. The first thing we must then do is to make certain that we know the former before we can reasonably undertake to apply it to the latter. From the very beginning therefore we are involved in a process of discovery, one that cannot be done by asking all the powers of the soul to reveal their secrets and participate as equal partners in the work to be done. Only the cognitive powers of the rational part can be trusted from beginning to end. Neither the *epithymêtikon* nor the *thymoeides* can be of any help. In fact by allowing them to play a role in this delicate mission, their interference becomes a meddling source of constant confusion. Since the practice and study of politics depend for clarity on the unencumbered function of the *logistikon*, not only do we rely on it for the correct direction to follow but equally so for the selection of the means for the soul's continuous development. The *aretê* of the soul is ultimately dependant on the proper tending of the reasoning power. But so are ultimately the excellences of the other two parts of the soul, no matter how self-centered and resistant they tend to be to the intrusions of rational controls.

The strategy is shaping up in clear outline. Once the *polis* in *logos* is constructed, the interlocutors will be free to dramatize cultural history to show how it pertains to injustice as a diverse and multiple image of political decline. But first we need the model. With that done, we will be ready to move from the *polis* in discourse to the "outside" *polis* and gain an overview of how the drama of political decline occurs in history. It is here that the criticism of the actual constitutions, especially democracy, will be viewed in proper perspective and comparative terms. In all cases, the main criterion of judgment will be the kind and extent of damage each constitution inflicts on the souls of the citizens by misplacing values and reversing the natural order of the powers of the soul. The malfunction is done at the expense of human reason, the *logistikon*, in its natural role of being the *hêgemonikon*. The core of the criticism in the case of democracy in particular is how democracy allows and promotes the reversal of the soul's powers and why.

## 11.2 The Dramatic Setting

When Plato wrote the *Phaedo*, sometime in his fifties between 385 and 380, the aftermath of Socrates' death was still reverberating in the minds of his followers. Plato, now a master of the principles of dialectic, commemorated the teacher's death after the return from his first visit to Sicily in 387. The background of the *Phaedo*, the antecedents, had already been prepared in the *Apology*, the *Crito*, the *Euthyphro* and the other early dialogues, also quite likely, the *Symposium*. The time had now come for what was to be one of his most moving and rich in ideas

dialogues, placed at the end of the life of an unforgettable teacher who was executed in 399.

With the theory of Forms settled in the *Phaedo*, the time was ripe for the writing of the *Republic*. Socrates was adamant in insisting on the place of philosophy in the life of the *polis*. It was Plato's turn to state it as a principle of life in dramatic time and display it as a great moment of the *persona* and as its own narrative while exhibiting the causes of the crisis Athens continued to experience in the aftermath of the war. The time had come to talk about political failure in the form of a philosophical drama.

Socrates is brought back in the *Republic* to relate a great moment in the philosophical quest of the just life. As expected, Socrates and his interlocutors embark on the momentous search for political justice to define *dikaïosynê* in civilized life, already a central theme in previous dialogues such as the *Protagoras*, the *Gorgias* and others, including the *Symposium*. The moment had come for the full exploration of the implications of the theory of Forms for political justice and the kind of life the citizens of a just *polis* deserve to pursue, especially those who are prepared to call themselves philosophers.

A comment is needed on the dramatic setting of the *Republic*, especially with regard to the dramatic date. Much has been written about this item, as Guthrie (1962–1981, vol. 5, p. 438) has noted. He tentatively settled for the year 421, while acknowledging that J. Adam (1963) had opted for the year 410, if not 409. I must explain why I think the dramatic date Adam proposed should be preferred.

The narrator in this dialogue is Socrates. We have no information about who the listeners to his story are as it begins to unfold as an extended dialogue within a monologue, ending with the myth of Er presumably for the benefit of the unnamed companions and beyond. We immediately become aware of the presence of Socrates and a few of his companions, and then of more persons who are added as the narration progresses. Notably, among them Plato's older brothers, Adeimantus and Glaucon, active participants of the discussion at the house of Cephalus. Setting the dramatic date in 421 would mean that Plato's two brothers had to be ten and eight years old respectively, hardly old enough to participate as active interlocutors. Plato could not have possibly overlooked so important a detail and allow the discrepancy to remain in this work by turning his brothers who were still alive at the time of the writing of the *Republic* into *meirakia*, immature boys. If his two brothers, as we may assume, read the *Republic* soon after it was made available they would have been more than surprised to hear that their brother had endowed them at the pre-adolescent age with such alertness and admirable conversational skills. Since the two brothers distinguished themselves in the battle of Megara in 409 (*Rep.* 368a), the setting quite likely intimates an occurrence at some time soon after that year. If we settle for this dramatic date, we should assume that by this time Cephalus was still living with his sons, Polemarchus and Lysias, in Piraeus.<sup>1</sup>

The date of Cephalus' death is uncertain. He was probably quite young when he settled in Athens as a metic in 450. Polemarchus (circa 450–404) was executed by the Thirty in 404. In 409 he was forty, but if the date of the *Republic* is pushed back



to 421 he would be in his late twenties. Aside from this adjustment of his age, he could not have been in Athens in 421 since he and his brother Lysias had migrated to Thurii as colonists in 430 and did not return to Piraeus until 415 or later. Lysias (445–380) was thirty-six in 409, old enough to have built his reputation as a rhetor. Except for Plato’s “extending” the life of Cephalus to justify his presence in the *Republic*, the ages of the other *personae* support 409 give or take as the dramatic date.

Accepting the year 410 or better 409 makes believable the role of Plato’s brothers, now in their early twenties and with strong interests in philosophy, and Socrates at the ripe age of sixty, close to his age in the dramatic setting of the *Phaedrus*, the *Timaeus*, the *Critias* and the *Euthydemus*. The year 409, when Plato was about fifteen, would perfectly serve to justify an imaginative setting in which Socrates “recalls” the conversation he narrates in the *Republic*. The reader willingly suspends disbelief to follow the unfolding of the events in the dramatic world of the philosophical myths.

It must be remembered that the conversation at the house of Cephalus gradually led to the need to examine the idea of justice and why, once it was done, the discussion moved on to the role of education. The issue was raised to show why after defining justice by way of constructing a model of a just *polis* and defining analogously the “just” citizen, the education of the future leaders had to be decided. Two principles stood out: the generic nature of the human soul and subsequent to it the recognition of the unequal distribution of competence which the powers of the individual souls possess. Since the aptitude of each soul in turn decides the place of the individual in the just city, the second principle concerns the analogy between just soul and just city. As there can only be one just *polis* the various souls are granted the opportunity to rise to their appropriate level of accomplishment by accepting their place in the well-being of the *polis*. Each type of soul contributes to the *polis* what it can do best. This is a condition that renders judiciously the role of each individual in the pattern of justice. The analogy of the just *polis* and the just soul is eventually clarified logically with the theory of Forms and illustrated mythically in the Allegory of the Cave.

Understanding the character of the just citizen calls for immediate attention to the required virtues: wisdom, courage, *sôphrosynê*, and finally justice. Granted this requirement, the question is how the set of virtues will be exemplified in the *polis* in discourse. If the *polis* has the virtues, it does so because its citizens have them. But, then, how does the just *polis* come to be given that there are so many and diverse individuals? This a problem that calls for knowledge of the nature of the soul, the faculties (*dynameis*) human beings possess by nature and how they need to be trained through a development program, the appropriate *paideia*. Given that the individual is not born virtuous, how is excellence of the soul acquired? The same holds for the *polis*. According to Plato, all depends on the role human beings assign to the power of the *logistikon*.

The upshot is that since we are not born just by nature, we can become just by nurture, a process that needs to be planned to guide all cultural endeavors. In view of the diversity of human aptitudes, the citizens are divided into classes or

groups, depending on which part of the soul is entrusted to make most of the practical decisions. The crucial problem here is what role does each class assign or can assign to the faculty of the *logistikon*. This assignment will determine the limits of the place of virtue and the efficacy of its function in each class within the collective life of the citizens. Leadership in the political structure, then, can only be decided with reference to which virtue or virtues become available to each class and to what extent the dominant role is entrusted to the *logistikon*. The political effectiveness of this *dynamis* depends on whether it will serve as an instrument or as a self-sufficient excellence. Without this assignment, the acquisition of knowledge and its attainable breadth will be limited to whatever rules the other powers dictate once they become the dominant parts of the soul. Such are the guidelines that will stamp the various types of government as well as determine the type of justice attainable in each. The just *polis*, in the unqualified sense, can only be one that recognizes the primacy of the *logistikon* to promote all the virtues and enforce an educational policy suitable to maintaining the class of just leaders. As the political pattern of the just *polis* becomes visible so does the justice in the individual (435d–444a). This attainment is rendered secure on the condition that the Form of Goodness, the *agathon*, is identified and recognized as the ultimate foundation of all conduct.

There is still the question of the practical possibility of the just *polis* in discourse.<sup>2</sup> If I understand Plato's view correctly, the *polis* in discourse serves its purpose only as a model. What is left is not the possibility of its realization as an actual *politeia*, transferable from *logos* to *politikê technê*, but a model of just conduct for all existing cities to emulate as far as their circumstances permit. The model only points to the perfectibility of human nature as a guide to identify the limits of service that the existing constitutions can offer to their constituents. Hence, as I will try to show in another section, the criticism of the diverse systems of government serves as a reminder of the compromise of the Form of Justice in the existing constitutions. While democracy is launched on the principles of equality, *isonomia* and *isotimia* and opens the gates to its citizens to choose their own way of life, the privileges it provides can easily be misused. Given the model of justice of the city in discourse, the difficult question is how far democracy can travel on the road to the *agathon* from initial freedom to the attainment of justice for all citizens. Furthermore, how does democracy provide the citizens with the opportunity to lead a philosophical life? As Guthrie phrases this problem, "The search is less for a city than for personal righteousness."<sup>3</sup>

All actual cities, should they decide to make use of the *polis* in discourse, will face the problem of persuading the citizens to accept guidance and act according the principle of reason, given the way human faculties function, the variety of talent and range of efficacy of the *epithymêtikon*, the *thymoeides* and the *logistikon*. Noting the frequent lapses from ordinary work into pathological conduct, where deficiency in *sôphrosynê* renders the failures most visible, helps us understand why injustice is a state into which human beings so easily backslide, often to the point of insanity (*mainomenos*).

### 11.3 *Eros Philosophos: From the Symposium to the Republic*

Given that the moving force in the individual to rise to the level of virtue is eros, when left to its own devices it becomes a compulsive force, more often than not incurable.<sup>4</sup> When this happens at the community level, the *polis* follows a dangerous path prone to dissolution and steady decline.

In dealing with the erotic element in the pursuit of knowledge, Plato is in effect making a transition from the *Symposium* to the *Republic*. Eros in the political life ranges from the lust for power to the love of the Good. To grasp the significance of Plato's position, we need to supplement the lessons of Diotima with the psychology of the *Republic*. At *Smp.* 209a–b Diotima tells Socrates: “by far the most important kind of wisdom is that which governs the ordering of society, and which goes by the names of justice and moderation.” After praising Lycurgus and Solon as wise law-givers who benefited their cities as well as humankind, she urged him to understand the final lesson about climbing the ladder of love (210a–212a): to contemplate “the beauty of laws and institutions” until from the beauty of every kind of knowledge “to turn his eyes toward the open sea of beauty” (210d). This last lesson will be expanded in the *Republic* to include the theory of Forms as the ultimate foundation of justice in political life, a theory which itself rests on the power of the soul to function under the guidance of reason.<sup>5</sup>

The argument in the case of the *Republic* is to show how the rungs on the ladder (*anabathmoi*) of erotic life in the *Symposium*, as the lessons of Diotima instruct and which Socrates confesses he did not quite fully understand, are finally grasped and applied to the *politika pragmata* in the *Republic*. We have in this dialogue the most complete treatment of the nature of the *philosophos*, whether in the form of the philosopher-king or the king who must be properly educated to become *philosophos* and as such help raise the *polis* to the level of justice. In either case, only the philosopher can save the *politeia*, any *politeia* that has lapsed into a deviant type of political rule. After the *Symposium* and the *Republic*, the next step to understand the full power of *eros philosophos* came with the writing of the *Phaedrus*.

*Rep.* 490a–b takes us clearly back to the *Symposium* and the ladder of love illustrating the way the philosopher rises to “real being” so that the pains to give birth in real beauty will no longer be felt once the work is completed. The philosopher, finally, “by consorting with real reality he would beget intelligence and truth (*gennêsas noun kai alêtheian migeis tô onti on*).”<sup>6</sup> And at *Rep.* 485a–b we read:

Well as we said to begin with, it is first necessary to understand their nature. I think if we once sufficiently agree on that, we'll also agree that it is possible for the same people to have both, and that no others should lead cities.

How so?

First let us agree that philosophical natures are always in love with a study which makes clear to them the nature and reality of what always is, and is not caused to wander by coming to be and passing away. (Allen transl.)

We read at *Apology* 29d–e what Socrates indeed considered his mission throughout his life. And in a memorable passage where he considers the chance of the jurors acquitting him on the condition that he stop philosophizing, he replies:

I hold you in friendship and regard, Gentlemen of Athens, but I shall obey the God rather than you, and while I have breath and am able I shall not cease to pursue wisdom or to exhort you, charging any of you I happen to meet in my accustomed manner: "You are the best of men, being an Athenian, citizen of a city honored for wisdom and power beyond all others. Are you then not ashamed to care for the getting of money and reputation, and public honor, while yet having no thought or concern for truth and understanding and the greatest possible excellence of your soul?" (Allen transl.)

The *Apology*, being Plato's first dialogue, has set dramatically a historical event for all the dialogues to confirm. What the mission came to mean is stated again in the *Symposium* in a phrase that places the philosopher between the wise and the ignorant (*philosophon de onta metaxu sophou kai amathous* 204b). As drama, this radical admission reflects the unceasing effort to stay the correct course in the quest for truth while living in a city where injustice has been allowed to become part of the Athenian democracy. The ironic praise in the quoted passage from the *Apology* was wasted on the majority of Socrates' jurors. Athens, the greatest and most famous city in the world will soon pass the death penalty on Socrates. Years later, when Plato will write the *Republic*, the greatest city will bear a deserving resemblance to the city of Pigs. The irony became a haunting theme: the greatest city, although not completely bereft of philosophy, was still lacking the determination to avail itself of the corrective wisdom to become just. Athens remained another city that had lost its soul (*hêmartêmenê politeia*).

## 11.4 The Criticism of Democracy

In this section I discuss two topics, the error Plato's critics commit, actually a case of anachronism, and Plato's critique of democracy as another type of *politeia* with an unjust constitution.

*The error of anachronism.* To begin with, we need to take into consideration the fact that democracy as it emerged in modern states is neither a replica nor a continuation of the democracy of the classical Greek version. The ancient and the modern conceptions of democracy are two different political entities that developed under different conditions in response to different social and economic problems, the partial overlapping values notwithstanding. The modern democracies rose as an answer to the abuse of power in oligarchies and autocracies, both remnants of medieval feudalism. As a movement, modern democracies sought to obtain individual rights in the name of freedom and protect the rising new class of property owners. The related concept of justice shows little connection with what Plato discussed as the search for the just individual under already existing conditions of freedom and established rights in the Athenian democracy. These political prerogatives had been established in a series of reforms going back to Solon, Cleisthenes and others. Not the acquisition of rights but the abuse of democratic rule was what concerned Plato. The concept of justice related to individual rights has little if any connection with what Plato discussed in his search for the just individual under the already existing conditions of the Athenian democracy. His objective was the correction of the aberrations

of democracy not how to establish one. What was admittedly missing at the opening of the conversation in *Rep.* I was the definition of justice beyond what the then current views advocated. In addition, Plato aimed at the improvement of the existing democracy by means of an educational program built on a rational definition of justice, one that could help the future citizens acquire the requisite virtues that grant the soul its harmony. To put it somewhat differently, Plato's aim was to redirect political conduct to secure the development of the human faculties under the guidance of reason. As such it announced a radical project in politics.

The problem for ancient democracy, as Plato saw it, was to secure the role of reason in the affairs of the citizens. This concern points to another fundamental difference between the role of political theory in classical democracy and its modern counterparts. The asymmetry, unless properly understood, can easily lead to misconceptions and condemnatory judgments. Whereas we depend rather heavily on the presumed sagacity of legislators to equalize the spreading of rights and privileges, Plato put his faith in the educational reforms that alone could assist in the development of the needed habits to help turn the powers of the soul into virtues with the proper unity.

By following this course of reasoning, I will try to show how Plato did not unqualifiedly condemn democracy, despite the abuses this form of government suffered and the calamities its leaders brought upon the Athenians during and after the Peloponnesian War. These aberrations include the death of Socrates following the restoration of the democratic rule in 404. At the heart of Plato's approach to define the needed principle of justice in political life, once we bypass the idea of immortality, is his theory of human nature, what is referred to as the soul and its powers. It is the soul that first and foremost is in need of justice. Education, next to understanding natural aptitudes, is the institution that provides the means that contribute to making the citizen just or unjust.

First we need to follow the construction of the model of justice. It has been said that Plato is no admirer of direct democracy. This is a wrong-headed inference. Even if the just *polis* were to come about, not all citizens could qualify to occupy the office of legislating or that of designing the educational curriculum. The Guardians, being an exclusive and privileged body, remain responsible for maintaining social harmony among the diverse classes just as they are for the defense of the *polis*. But no ordinary citizen is in principle excluded from ruling if it can be shown that he/she possesses the requisite native qualifications. If an individual is selected to receive the demanding training for a place in the body of Guardians and meets the demands of the arduous task, the door is open to the responsibilities of leadership. Plato takes for granted the natural fact of the inequality of initial talent, although all souls formally share the same nature.

Human beings have the same set of powers but not the same endowment. This for Plato is a fact of nature, one that makes human beings at once equal and unequal. Since attaining justice of the soul also remains a social demand, the individual limitations dictate the level of attainable virtuous development and political conduct for the participation in the welfare of the *polis*. These considerations circumscribe the limits of democracy and the level of participation in the shaping of public

institutions. To ignore them is to court disaster just as to misunderstand individual differences is to invite more unjust acts than what the *politeia* can afford before lapsing into a worse type of deviancy. On the whole, so it seems, Plato is not an enemy of democracy, but rather an austere critic and teacher, what Socrates called himself in the *Apology*.

Plato's position is clear. If Plato was really concerned about philosophy as a way of life to be made available to all citizens and all human beings, since they are by nature entitled to gain access to philosophy, then in democracy this goal is made initially available to every citizen. It is understood, however, that as a constitution democracy will be judged on the basis of whether it can secure the proper conditions of *isonomia*, *paideia*, and *dikaiosynê* for the *enaretos bios*.

In order to grasp more adequately Plato's critique of democracy, as well as the other types of government, we would need to go back to the opening scene of *Republic* Book I. The main theme that will dominate the entire work comes to the foreground as the concept of justice (*dikaiosynê*). In a preliminary way, it should be said that Book I is definitely an organic part of the entire work and not an addendum written earlier and only later attached to the *Republic* as an afterthought. I feel compelled to stress this point because in my opinion the discussion on justice does not get underway as preliminary to the theory of Forms but rather as a pressing political principle frequently distorted and abused. This dislocation of justice happens even when intentions are sincere. Plato correctly sets the dramatic pace by having the *personae* showing at the opening commendable religious tolerance towards imported new types of celebration, as in the case of Bendis. Such doctrines of belief are not seen as threatening to the established orthodoxy since there was none.

Once the return to Athens has started, Socrates and his companions will change course after accepting the invitation of Polemarchus, son of Cephalus, to gather at his father's home. Once there, the companions and other guests will embark on an exploration of the place of justice in life. There are strangers and non-Athenians present, each with his own understanding of the meaning of justice. In a way, the discussion starts out "empirically," as a batch of opinions that throw into the open the discussion of justice.

Criteria of correctness are soon introduced as means to settle disagreements; they provide Socrates with the opening he needs to initiate the elenchus. One by one the proposed views of justice will be found wanting in correctness until they are shown to lead to an impasse. At that point Adeimantus and Glaucon press Socrates to come up with a definition of justice. Socrates accepts the challenge but is quick to admit that he cannot deliver a direct answer. This admission is exactly what brings in the indirect approach: the need to build a model, namely the just *polis* in discourse. The implication of the new direction is clear: the realm of opinion in all its diversity is no better than the existing types of government, all of which claim to serve justice but hardly promote this goal. The consequences of Socrates' procedure will haunt the interlocutors to the end of the *Republic*. Stating the same in a more acute way, *logos* embarks on the relentless critique of all existing forms of government that claim to be agents of justice.

There are two versions of democracy in the *Republic*. There is first the actual democracy as practiced in the *polis* of Athens and elsewhere, and then there is the pattern of democracy depicted as a constitution once the model of the *polis* in discourse is put aside. The two types are distinct and unequal. The difference between these two versions should alert the reader on how to avoid transferring back and forth critical reservations derived either from dissatisfactions with the actual democracy or from defects noticed in the analysis of the shortcomings of what are shown to be deviations from the “good” city of aristocracy. But there is more to this warning and it has to do with the way many readers of Plato have tried to understand the *Republic*. The dramatic principle on which the entire *Republic* is based, a story within a story spinning actions and ideas centered on the device of a *polis* in discourse, makes the extrapolation of formal principles to pertain to historical situations a distortion of the intent of Plato’s dramaturgy. I think it is theoretically incorrect to use the defects of the actual democracies to criticize Plato’s type of democracy, actually an aberration of a “good” constitution, as though the two distinct forms of democracy have the same origin and serve the same goals. The two types should not be treated as though they are members of the same political species. Making the free transference from one to the other leads to distorting Plato’s argument and hence to misinterpreting the *Republic*. Regrettably, there have been too many misleading approaches and readings of Plato’s position. Democracy as a deviant constitution is treated in the *Republic* mainly in light of the attainability of justice as understood in the model Socrates has set up in the quest for justice, individual and political, as presented in the constitution of aristocracy. Of course, there is no denying that historically the development of democracy as a form of government was instituted to remove grave malfunctions of legal institutions. Actually, the operating principle was the introduction of a measure of *isonomia* and *isotimia*, but whether either of the two desired conditions actually led to a corresponding practice of justice is another story. The abuses of power show that democracy, like the other deviant forms of government, led to social inequalities and dissatisfactions. *Isotimia*, to the extent that it was enforced, proved to be a necessary but not a sufficient condition of justice. Nevertheless, the presence of *isotimia* gives this actual type a special status. The issue, however, is whether this form of government can prevent the occurrence of injustice, let alone remove the ensuing evils. This is the basis for Plato’s criticism, and it is ultimately best understood in the context of the just *polis* in discourse. Consider, on the other hand, the cases of oligarchy and monarchy, where injustice tends to be more rampant and destructive.

The criticism of democracy is meant to propose ways that would be corrective of existing conditions in this type of government. But the proper way of life, i.e. the securing of the conditions for the philosophical life as a constant pursuit of wisdom to attain *aretê*, presupposes the clarification of what it means to live as a *philosophos*. This is the reason why the *Republic* as a dialogue aims at defining the *dikaïos anthrôpos*, and this in turn requires the concept of *philosophos*. Ultimately, the philosophical drama of this dialogue is about *philosophia* itself. After writing the *Phaedo*, Plato had to return to the unfinished business of the good life in the *polis*. That is why he wrote the *Republic*. As Socrates states at *Rep.* 500c, “Familiarity



with the divine and orderly makes the philosopher divine and orderly so far as a man can be.” The quotation calls for a comment on the significance of the pursuit of excellence embodied in what piety regards as the divine and in what ways it is available in human conduct. Reaching for the divine sets a standard for all virtues but as a pursuit it does not make the divine either totally remote or irrelevant to the human condition. Given that the standard has political overtones, Plato proceeds to advocate the aspiring for the divine as a right of all citizens. A constitution that either distorts the concept of the divine or limits its availability to few citizens fails to function properly and makes the place of philosophy in life a secondary good, if a good at all.

It is not an accident that the opening of the *Republic* places a special value on the rites of the Thracian religious rites of Bendis as an occasion for celebration, or that the Panathenaia provides the background for the *Timaeus*. However, it should also be said that Plato’s idea of religion should not be confused with religious practices in other traditions. The Greek polytheism was more tolerant and open to divinities other than the official ones in the city, for example of Athens. The receptivity of celebratory occasions was part of the public conduct. There was no official state clergy or official state dogma. The religious celebrations belonged to the city, not the opposite. The uses of religion in public life were for the Greeks a delicate matter. Socrates’ trial that brought to the surface accusations of religious misconduct was not a simple matter stemming from either atheism or irreligiosity. But that is an issue that calls for another special study that cannot be undertaken here. What interests us in the main is how Plato assigns a proper place to the divine in the life of the *polis* as it relates to the pursuit of philosophy as both a way of life and as a political corrective.

## 11.5 Democracy and the Decline of Constitutions

Before we proceed to discuss in more detail Plato’s reasons for criticizing democracy, we need to understand what he means by referring to this political system as lacking in power to promote justice. Lurking in the background are many questions: How was democracy practically in force during Plato’s time and in earlier decades, how did it come into power and why, what were the ends it was expected to promote and whether it managed to serve them well, and how and why did it fail its constituents as well as its leaders? What we read in Plato’s *Republic* is an account of the causes of deviation from an originally just, hence good form of government, be it aristocracy or kingship.

Since democracy is one of the declining political forms in a series, ranging in descending order from aristocracy, timocracy, oligarchy, democracy and finally to tyranny, we need to identify the process of degeneration from one imperfect type to the next. As it turns out, Plato is careful to admit that even the government by aristocracy, once functioning in practice, cannot effectively prevent its own deterioration thereby yielding its place to the next in line (*Rep.* 543a–576b). The key to understanding the entire process of decline is found in the differentiation between

the just *polis* in discourse and the good *polis* as an initial constitution in time. Which of the two constitutes the “starting point”?

Ultimately, the cause for the degeneration must be sought in the characters of the citizens. Once certain flaws are allowed to remain undetected or, even worse, accepted as normal, they inevitably generate actions leading from bad to worse. At first sight, the flawed characters, understood as causes, lead one to conclude that a pessimistic inevitability of decline haunts the human race. This is one way of reading Plato, though not the most constructive. The problem becomes more serious when we stop to think about reconciling the stability of justice in the good *polis* in relation to the beginning of the failure of justice all the way to tyrannical rule. Briefly put, the diagnosis shows that the beginning of political failure in justice is due to no other cause than the failure of the citizens to pursue the appropriate state of virtue. The issue of the degeneration of types of government remains one of the most puzzling features of human existence. One thing must be made clear at this point of our discussion, and that is whether the answer to the question of degeneration is caused by internal imperfections of the “just *polis* in *logos*” itself. If that were the case, then the blame should be placed squarely on the shoulders of the Guardian philosophers who were unable to prevent the decline of justice despite their knowledge and meticulous planning.

Guthrie (1962–1981, vol. 4, p. 528) notes: “The genetic sequence is not meant to be taken literally, as is obvious from the fact that the Platonic state does not exist as a starting point.” The reading is well taken but we still need an answer to the question “What is the starting point?”<sup>7</sup> To state the matter differently, the starting point is something like a mystery. If I read Plato correctly, the starting point could not have been the just *polis* in discourse. As such, this logical construct never existed nor could it possibly ever come into being in the sensible world inhabited by the embodied souls of human beings. Admittedly, human beings as citizens are capable of development with chances of having some success but mostly prone to failures. The necessity of the *epithymêtikon* muddles the actions for the rest of the soul as it forces its needs to the foreground of action and enlists the assistance of the other powers, especially the power of reason. On this basis, then, we must not involve the *polis* written in discourse in what happens to any form of government, including the “best,” i.e. aristocracy, since even this form takes place in the sensible world of change. Furthermore, it needs to be remembered that even the *polis* in discourse is an image of the Form of perfect justice and as such, Plato tells us, it can only exist in heaven with all the other perfections, although as a logical construct it is made in *logos* by human beings, the interlocutors in the *Republic*.

We are left with one alternative: to think of what has been or could have been a good *polis*, given the conditions of our sensible world as the temporary habitat of our ensouled bodies or rather incarnated souls. Such a good *polis* would thus be the starting point of the long process of political deviance as well as the source from which to derive applicable criteria to diagnose the flaws that will unavoidably plague the subsequent deviant governments.

Hence the good *polis* is introduced against the variations of which the poets had spoken of in their myths as marking the beginning of the civilized life. Pictured as

the golden age of humanity and understood as the starting point of political devolution, it amounts to nothing more than an image of the *polis* in *logos*, itself an image of the Form of the Just and the Good. The starting point is therefore an image of an image, twice removed from true being. If this is acceptable as a correct application of Plato's theory of knowledge as it pertains to the construction of the just *polis* in the *Republic*, it follows that the criteria for judging any deviant constitution must derive their pertinence from an initial historico-mythical *polis* that conceivably may have existed at some point in the remote historical past. It is this image of an image that provides the basis for criticism in political conduct, as the judging harkens back to that noble beginning. When we turn to Plato's political vision, we see how the philosophers in the *polis* in *logos* in the *Republic* accepted the assignment Socrates and his friends granted them on the basis of the special dynamics of their exceptionally gifted souls. The gifted philosophers, just as in a masterful work of art, were pleased to accept the assignment and devote themselves to its demanding tasks. From that point on, the philosopher-leaders rule in justice unmistakably and unselfishly. The *polis* in discourse remains what it is: a model and an image of the Form of Justice. It has earned its closeness to grace and beauty in eternity. As such it neither exists in the sensible world nor does it ever become a starting point in the adventures of the political affairs of mortal life. Whereas the philosopher-leaders in the *polis* in discourse perform in perfection, the *aristoi* leaders of the *polis* in mythical history remain subject to the uncertainties of the dynamics of the untrained faculties of the soul of their subjects and open thereby to the pressures of the desires that more often than not reach out to grasp for whatever images of virtue become available as reason struggles to sustain the wavering *eros* for an elusive justice.

Books VII and VIII provide us with the discussion to understand the *metabasis* or transition from the ontic base of the realm of the Forms to the realm of life on earth. The interlocutors had not forgotten that the original quest was to define the just person, what that person can be like in the sensible world. If Socrates was already convinced that it is impossible for such a person to exist on earth, one wonders why the discussion was undertaken at all. In the course of the quest, he had to move the discussion in the direction of the best-suited highest plateau of human reason: the Forms. Aware that the Forms exist in only a distinct and separate realm, beyond the world of change, the enterprise was entrusted to the power of dialectic. Once the Forms of Justice and the Good were shown to hold the secret of the permanence of values, defining the just person became an easier task. Conforming to the Forms as the end of the philosophical life in the pursuit of wisdom provided the criteria for judging individual and collective conduct. After the Divided Line was understood and the limits of knowledge set, all that could be done to define the just person via the Form of Justice had come to an end. The next move was to return to the conditions of life in the sensible world, to the problems of political justice. Thus the conception of the good and just person as understood through the perfected model of the *polis* in discourse, had to be transposed through a *metabasis* to the *polis* in the sensible realm. The aristocracy, as pictured in the *polis* in discourse, had to be shown as having its counterpart in sensible reality, for otherwise all that went on before

would lose its relevance to the conduct of ensouled beings. Socrates turned to the only good constitution on earth: *aristocratia*, be it a government by one *aristos*, the case of kingship, or the many *aristoi*. All other alternatives could only be deviations. At 544e we read: “— Then, if there are five kinds of city, there would also be five fixed dispositions of soul among private citizens. —Certainly. —But we’ve already described the man who is like aristocracy, when we rightly claim is good and just.”<sup>8</sup>

The passage from aristocracy to democracy through the intermitting states of timocracy and oligarchy continues the intensified unruliness of the cravings of the appetites and the outbursts and excitements of the *thymos*, the spirited will. When the abuse of power in the performance of oligarchic rule reaches the point of no return and justice falls into a state of disarray, democracy becomes an unavoidable solution to the growing problems of social dysfunction. It rises as the next stage of political life aspiring to recover justice with the newborn freedom and the distribution of power through direct participation in the ruling of common affairs. Eventually, the individuals as citizens, the *demos*, rulers and ruled, fall victims of the unruly state of their souls. The appetites once more control the *thymos* which in turn overpowers reason and dictates its course.

Plato has given us a powerful description of the processes that condition the transition from one type of government to the other, down to the last resort in tyranny. No need to repeat here what Plato has done so eloquently, even if the causes of the transitions are not uniformly applicable to all states in the history of political life. My concern has been more with Plato’s critique of the traits that account for the continuation of injustice, political and individual, once democracy replaces its predecessors. Why and how democracy brings with it the seed of injustice that will in the long run lead to its own demise and to another kind of injustice as in the tyrannical rule, is another topic and has been explored by historians and well as by philosophers many times. Without going into a detailed description of the state of affairs as developed in democracy, the end result takes either of two roads: a discord in personal conduct or disharmony in the public institutions, in both cases generating innumerable instances of injustice. They stand for situations at once exploitable and decrepit, both disruptive of the effort to bring about excellence in character and rational agreement (*homonoiia*) in the *polis*.

Viewed from the other end of the spectrum, democracy brings with its statutes the promise for the greatest amount of pleasures to the greatest number of individuals. This is no minor benefit and cannot easily be discarded unless different criteria of evaluation of the ensued satisfactions are introduced.<sup>9</sup> Plato knows what democracy can do by way of gratification of the *epithymêtikon* just as he diagnoses the detrimental effects it can have on the pursuit of justice and the cultivation of reason.

The unjust democratic citizen harbors a potential tyrant because while the *logistikon* is compromised and turned into a slave, both the *thymoeides* and the *epithymêtikon* are given free rein to decide what action is best for the gratification of the unclear demands of *eros*, be it *eros* for power, fame, or sexual pleasure. Eventually all three faculties indulge in injustice and contribute to its spread over the entire *polis*. On the other hand, only in democracy can the citizens enjoy free reign of the soul’s functions, but do so at a formidable price and at the expense of

virtue. Nevertheless, the fact is that only democracy can turn the *politeia* into a field of novel opportunities to test the display of the soul's powers at a rapid and general way as a universal prerogative sanctioned by law. The stakes are high and the risk harbors unpredictable consequences. No other constitution can put the human soul to the test of justice and expose it to the dangers of expansive injustice as democracy can. There is one thing that augurs darkly for the days ahead in democracy: how far the trappings of the *epithymêtikon* will enslave the only means available to human beings to steer the course of political action in the direction of justice. Should the move toward this and the other virtues succeed, democracy would cause the withdrawal of democracy and point the way back to aristocracy.

This paradox of the democratic rule stems directly from its status as being a *meson*. If democracy moves decisively in the direction of *logos*, it will eventually repudiate the unconditional *isotimia* and proceed to transform itself into an aristocracy, an initially good state, by keeping only certain indispensable elements of legal protection for the welfare of the citizens. If however it should move in the opposite direction and unleash all the daemons of the *epithymêtikon*, sooner or later the pressure of the *thymoeides* will drag the *polis* away from liberty and force it to submit to the extreme of tyranny. The best and the worst move darkly in the secret folds of democracy, a tension that explains the joy of initial victory as well as the deep pain over its potential failure. The worst feature of democracy is its potential to breed the future tyrant, the *tyrannikos anêr* (568d7, 573e7, 574a1, 575a9–b9). The suffering that the abrupt and violent coming of democracy causes, when it succeeds through revolutionary action, and the damages that come after its demise are immense, mainly because vast portions of the population become involved in the course of unruly changes. But the worst happens after the breakdown of the democratic rule. The *thymoeides* of the emerging tyrant reaches its highest expression at the moment of triumph as it tries to consolidate its newly acquired power by repressing all liberties as the cause of criminal desires (554b7–e2). The *logistikon* is now in the grip of the unbending *thymos* serving the new master with its unrelenting determination to save the *polis* from all its weaknesses and against its will.

Perhaps the main problem of democracy, stated negatively, is its inability to prevent the collapse of its institutions and avoid its down sliding into the perils of the tyrannical rule. The only course of self-preservation a democracy can follow is that of the collective intelligence of the citizens to legislate wisely, a type of action, however, that presupposes the possession of sufficient wisdom to initiate just policies. What this means is that with the coming of democracy the taming of the desires and the promotion of moderation are supposedly already in place. But given the free reign of liberties without corresponding constraints, the effectiveness of controls will always be in a state of coerced balance rather than one of *harmonia* as freely chosen *sôphrosynê*. Hence the attainment of virtue in democracy is always problematic and perplexing, actually becoming a plaything of *doxa* rather than a commitment of reason. The trouble with education in democracy is its ambivalence towards the role of reason. As *paideia*, it tends to give priority to the acquisition of skills, from survival devices to demagogic rhetoric such as refined public figures

and clever consumers employ as they hunt for available pleasures and honors. As devices, more often than not, they are dressed as superior values and distributed widely in the name of *isotimia*. In the meanwhile, the production of wise citizens is left to fortune.

The democracy Plato discussed in his *Republic* and other dialogues projected the problems that haunted the Athenian version, the city where the crisis of an over prolonged war had brought in its wake *hubris*, anarchy and shamelessness (*hybris*, *anarchia*, *anaideia*, 560d 4–5). Plato, writing the *Republic* some twenty years after the ruthless oligarchic rule of 404, had strong memories of the events that transpired during that period. Assuming the dramatic date of the *Republic* to be 409 when it opens, Plato knew exactly what thoughts and sentiments the drama required for its effectiveness on the readers. The setting employed strong ingredients from the practice of democracy prior to as well as after the oligarchic rule. With the return of democracy the citizens accepted as normal the unreflective chasing of pleasures and self-instituted practices. At this point of decline, the educational programs as elaborated in the *Republic* were at once undesirable and irrelevant.

The common good, the *koinon agathon*, when overshadowed by the private interests as demanded from the perspective of the *epithymêtikon* and abetted by the tenacity of the *thymoeides*, is no longer open to the corrective role of the *logistikon*. Freedom had become a popular idea affecting the promulgation of laws to guarantee the privileges gained, even to treat the *politeia* as another *idion*, another personal good. What replaced the *koinon agathon* in the long run, Plato concluded, was the trend to make the *polis* itself an *idion* available to all citizens. Competitiveness and greed do the rest except when the laws, already in the hands of the most powerful citizens, place restrictions and penalties to control excessive acquisitiveness. The philosopher becomes a marginal thinker easily replaced by the democratic Sophist. Such has been the dramatization of the compromise of the *logistikon* in the political life in the *Republic*. At the other end of the spectrum, if reason is given the opportunity to function openly and unencumbered, it will find its proper end and bring about the full understanding of the Forms as the foundation of political justice. As Plato was tirelessly warning his fellow citizens, the causes of political failure and the inability of humans to prevent the evils of injustice are rooted in history, not in the *polis* in discourse. In the *Seventh Epistle* he will repeat what he said in the *Republic*: either the rulers must learn to philosophize or the philosophers must become the rulers.

The Platonic account of the decline of constitutions from aristocracy to tyranny provides the proper background for the reservations regarding the ability of democracy to promote and secure justice. The focal point of Plato's account is that the decline begins with early signs of weakness in the aristocratic rule. It is, as already mentioned, the starting point. Aristocracy "in history" as a good state in the *Republic* (543c) whether ruled by one or few already contains the seeds of corruption. To be exact, the same seeds are in the souls of the people, independently of the soul(s) of the ruler(s). The same weaknesses will reappear *en masse* after the constitutional decline descends to democracy and liberty is widely made available. Gradually injustice becomes widespread.

The good constitution, as the starting point of political decline, resembles in part the just *polis* in *logos*, primarily because of the presence of philosophers in both. The just *polis* in *logos*, however, being at once perfect and just, remains unrealizable in the world of history and as such it is not subject to corruption. The critics of democracy, therefore, are the interlocutors in the *Republic* as they use their logical powers to assess the affairs of citizens in the constitutions that fail to meet the needs of the human soul in its entirety. Hence it is fair to say that Plato's critique of all constitutions is rooted in his theory of human nature and not the perfect status of the perfectly just *polis* in *logos*.

When democracy comes in for critical evaluation, Plato is sure to see it as another constitution that fails to help the soul attain excellence in all of its parts. The flaw in this constitution is shown to be failure of intelligence in action and a triumph of unbridled lust after pleasure wedded to greed for power, both harnessing reason to their service. *Eros*, crippled and confused, has lost sight of the ladder that invited climbing to Beauty.

Given the condition of generalized liberty in democracy, new versions of injustice continue to multiply until tyranny becomes inevitable. What started as a fair promise to remove the evils of oligarchic rule prepared the stage for the worst constitution: tyranny.

It would be unfair to Plato to call him a pessimist or even a perfectionist. When all is told, Plato kept the door open for the recovery of confidence in philosophy to attain the harmony of the soul even in the most unjust constitution. This point was stressed early enough in the *Apology*. Political justice, to the extent that it can be realized in any constitution, like philosophy itself, is for the individual to pursue, not possess. But without the cultivation of reason neither the individual nor the constitution can avoid falling prey in the snares of injustice. Justice therefore will always remain a human achievement, not a gift from heavens, and as such it will show itself as the measure of human sanity.

## 11.6 A Postscript

One wonders why after the *Republic* was made available the leaders in the democracy of Athens learned nothing from Plato's criticism, especially his diagnosis of the causes of war. When Plato died in 347, he left behind a philosophical school and students to carry on with the political vision he bequeathed. Some of the students in the Academy went on to advise their governments, Aristotle became briefly the teacher of Alexander with not much success, but accomplished far more as a philosopher in his own right. He wrote the *Politics* and the *Nicomachean Ethics*. His work continued and extended the Platonic concern for the life of the *polis*. Others in the Academy, who took over the leadership of the school, chose to follow the tradition of Pythagoras. The same cannot be said for two other new approaches to political life. Epicurus, an Athenian, disappointed and disillusioned withdrew from active participation in the affairs of the *polis* to find peace in his Garden, the city within



the city. At the other extreme, Zeno, a metic from Citium in Cyprus, enlarged the conceptual picture of the *polis* only to make it a habitat with no physical center, the *cosmopolis*. Both were signs that the *polis* had lost its vitality and original *modus vivendi*.

Plato was fortunate not to live long enough to see the actual decline of his Athens and the marginalization of philosophy in its practical affairs. The coming of the Macedonian rule marked the transforming of Athens into a state struggling to salvage the remains of its independence. The Roman conquest finally completed a process that turned the classical *polis* into a relic of the past, glorious but still a glorious relic.

## Notes

1. See Nails (2002, pp. 84, 190, 251, esp. pp. 324–6) on the dispute regarding the dramatic date of the *Republic*.
2. On this, see also Guthrie (1962–1981, vol. 4, pp. 483–6, 483 n. 3).
3. Guthrie (1962–1981, vol. 4, p. 486); also, *ibid.* n. 2 referring to Jaeger (1943, vol. 2, pp. 347–57, “The State Within Us”).
4. Compare Guthrie (1962–1981, vol. 4, p. 478).
5. Santas (1988, pp. 43–7) has shown how the theory of *eros* in the *Symposium* anticipates “another famous ladder, the Divided Line in the *Republic* in which the Form of the Good is at the top” (p. 41); see also Santas (1980).
6. Comparable passages in *Smp.* 204b, *Phdr.* 249b, *Phil.* 58a.
7. Compare Nettleship (1963, p. 300): “The account of these various stages of decline begins with the fall of the ideal state.” However, he also sees the order of decline as giving the appearance of being historical whereas upon a closer look “the order of arrangement is logical and psychological” (pp. 294–5).
8. The meaning of aristocracy contrasts with that of the rule of the *aristoi* in the *polis* in *logos*. This is clear from what Socrates says at *Rep.* 545c–d: “Come, then, I replied, let us tell how timocracy might arise from aristocracy. Is it simply this: that every constitutional change derives from the ruling class itself, when faction arises within it? If it is of one mind, even if it is very small, the constitution cannot be moved.”
9. Arends (1997, p. 22) notes: “It is especially by his viewing below the surface that Plato’s criticism of democracy becomes interesting for anyone who is interested in maintaining democracy.” Perhaps, but many reforms and preventive actions are required, and when properly instituted could conceivably lead the *polis* back to aristocracy, a process that in itself counters the continuation of democracy. Plato insists at *Rep.* 572d, that democracy is a *meson* and as such it could lead to either contrary. As Plato foresees the result, democracy yields more easily its place to tyranny.

## Bibliography

- Adam, J. 1963. *The Republic of Plato* (2nd ed.). Cambridge: Cambridge University Press.
- Allen, R., ed., trans. 1984. *The Dialogues of Plato* (Vol. 1). New Haven: Yale University Press.
- Allen, R., ed., trans. 2006. *Plato: The Republic*. New Haven: Yale University Press.
- Arends, J.F.M. 1997. “Could Plato Be Used as an Ally Against the Decay of Democracy?” In *Proceedings of the International Conference on Greek Philosophy*, edited by K.J. Boudouris, Vol. 1, 21–29.

- Guthrie, W.K.C. 1962–1981. *A History of Greek Philosophy* (Vols. 1–6). Cambridge: Cambridge University Press.
- Hamilton, R., and H. Cairns, eds. 1961. *The Collected Dialogues of Plato*. New York: Pantheon Books.
- Jaeger, W. 1945. *Paideia: The Ideals of Greek Culture* (Vols. 1–3). New York: Oxford University Press.
- Nails, D. 2002. *The People of Plato*. Indianapolis: Hackett Publishing Co.
- Nettleship, R.L. 1898/1963. *Lectures on the Republic of Plato*. London: Macmillan.
- Santas, G. 1979. *Socrates: Philosophy in Plato's Early Dialogues*. London: Routledge & Kegan Paul.
- Santas, G. 1980. "Plato on Love, Beauty and the Good." In *The Greeks and the Good Life*, edited by D.J. Depew, 33–68. Fullerton: Publication of California State University, Fullerton.
- Santas, G. 1988. *Plato and Freud: Two Theories of Love*. Oxford: Basil Blackwell.

## Chapter 12

# Function, Ability and Desire in Plato's *Republic*

Antonis Coumoundouros and Ronald Polansky

*Let us consider some object that is manufactured, for example a book or a paper-cutter; here is an object which has been made by an artisan whose inspiration came from a concept. He referred to the concept of what a paper-cutter is and likewise to a known method of production, which is part of the concept. . . . Thus, the paper cutter is at once an object produced in a certain way and, on the other hand, one having a specific function. . . . If man, as the existentialist conceives him, is indefinable, it is because at first he is nothing. Only afterward will he be something, and he himself will have made what he will be.*  
—Jean-Paul Sartre

*From this equality of ability, ariseth equality of hope in the attaining of our ends. And therefore if any two men desire the same thing, which nevertheless they cannot both enjoy, they become enemies.*  
—Thomas Hobbes

In his treatment of Plato's *Republic*, Gerasimos Santas emphasizes that there are two ways Plato approaches the good: "good as well functioning and good as perfect form" (2001, p. 58). Through much of the *Republic* Plato seems to appeal to the function of what he discusses to determine what that thing's good is. For anything with a function, its good should be performing its function well. Santas' focus on the role of function in the *Republic* is illuminating, and it is our purpose to expand upon his analysis.

In fact we believe that Santas quite correctly points to the significance of argumentation about function in the *Republic*, but he does not take this far enough. This is due to his supposition that there is an additional theory of goodness in the dialogue based on the theory of Forms, and the Idea of the Good in particular. In *Goodness and Justice: Plato, Aristotle, and the Moderns*, Santas states,

---

A. Coumoundouros (✉)  
Adrian College, Adrian, MI, USA  
e-mail: acoumoundouros@adrian.edu

there is no commitment to the idea that everything has a function, so there may be important limits to the scope of the theory; if there are things which can be good but have no function, the functional theory cannot account for their goodness. This turns out to be an important point when we consider the metaphysical theory of the form of the good. Platonic forms have no functions in the relevant sense. (p. 69)

But while Santas denies that the Forms have functions in the relevant sense, i.e., like things that exist in time, one may surely discover such functions for the Forms in relation to temporal things. If each Form serves as cause for a specific set of things by participation, then Forms clearly have definite functions. Whenever in the dialogues we have the suggestion that  $x$ ,  $y$ , and  $z$  are  $F$  due to  $F$ -ness, as all things are good due to the Good itself, the Form is surely performing a function of causing and accounting for the characteristic its participants share.<sup>1</sup> This is supported by the argument against Cebes at the end of the *Phaedo*. There without using functional talk, Socrates has the Form and then some other things explain a characteristic such as heat, cold, oddness, and evenness. For example, the Form the Hot itself can account for heat or fire can account for heat. Of most interest to us is that Life itself and the soul can be used to explain life (105b and 106d). But of course in *Republic* I 353d, life is just one of the functions assigned to soul. Hence it seems clear enough that if soul has a function that it shares with the Form Life itself, then the Form can have a function. Wherever something is serving as cause for something, it seems very possible to speak of the cause as functioning and having a function. We can, of course, be somewhat more restrictive, in accord with Socrates in *Republic* I: where what serves as cause can alone so serve or serves best, the cause has a function.

We have questioned that anything beyond function-based argument, conceived broadly, is ever needed by Plato to talk about goodness. Or perhaps we should say that whatever is good for Plato is so in connection with function.<sup>2</sup> Even the ultimate human good, happiness, is good functioning of the human, and other human goods contribute in various ways to this ultimate human good. Of course there are goods that are not strictly human goods, but these can be seen to be good in connection with non-human functions. What we first wish to consider in more detail, after a brief discussion of function, is how for Plato function connects with various abilities or capacities to perform the function. We shall show that Plato typically has a tripartite scheme of abilities that underlie functional competence. Then following our inquiry into the link of function and capacity, we reflect upon how Plato relates the abilities pertaining to functions to human desires. We argue that, for Plato, human desires are generally driven and maintained by the possession, or the perception of the possession, of various abilities.

## 12.1 A Brief Consideration of Function

Socrates uses function (*ergon*) prominently in his third argument against Thrasymachus in the end of book I (352d9–354a4).<sup>3</sup> Socrates suggests that different kinds of things from different classes have functions: there is a function of a horse,

of an eye, of an ear, of a knife, of the (human) soul. He affirms, "The function of each is what it can alone do or do best" (*hekastou eiê ergon ho an hê monon ti hê kallista tôn allôn apergazêtai*, 353a10–11, Shorey transl.).<sup>4</sup> "Each" here is ambiguous, since it can mean each *kind* of thing or each thing. Whether a class of things or a particular thing is at issue, function is determined by comparison with other classes of things or other instances of the class of things. Some of these may have closely resembling functions, e.g., a pruning knife and a carving knife, whereas others have quite different functions. To determine the functionality of kinds we compare kinds and to determine the functionality of individuals we compare individuals of the same kind. What something *alone* can do typically means not that this is the only thing that it can possibly do, but what it can do that other things cannot; and what the thing does *best* is what it surpasses all other things in doing rather than merely what it does better than the other things that it itself can do.<sup>5</sup> Hence, every member of a kind need not have a function if it does everything poorly and nothing uniquely, though each natural kind or class presumably has a function. And something or some kind may have more than one function if it alone can do several things best or uniquely.

Socrates continues the argument by pointing out that where there is a function there is an excellence at it. Virtue is what enables the function to be performed well or excellently and vice is the opposite (353b2–c2). So virtue is always in relation to function. The (human) soul, for example, has proper functions such as "taking care of things, ruling, deliberating, and the like," along with causing living (353d4–10). Socrates concludes the argument by saying that since there is a virtue of soul, the presence of this virtue, i.e., justice, enables the soul to perform its function well. So justice enables the soul to live well, i.e., to be happy. Regardless of the merits of this argument, we can see that Socrates links function to virtue and happiness.<sup>6</sup>

This use of function is meant to support eudaemonism, that is, the view that the main project of ethics is to determine the ultimate end and good for humans, and then to see how other goods contribute to this. Rationality in human action seems to Socrates to require an ultimate end or project in terms of which any other goods or projects can be evaluated. This is a teleological, and perfectionist ethical approach.<sup>7</sup> It suggests that function comes into play within a teleological, or means-end, context. Some things with functions, whether they are artifacts or natural beings, are serving as means to ends beyond themselves. Thus their functioning is not simply to their own advantage. The tool is serving the craftsman rather than the tool itself, and domesticated plants and animals may seem to serve their human masters even more than themselves. And similarly functional parts, as human organs, such as eyes, ears, arms, legs, contribute to a function beyond themselves, i.e., the overall operation of the living human being. Nonetheless, the good for that which has a function would seem to be performing its function well through having the necessary virtue for such performance of the function.

## 12.2 Ability in Relation to Function

In the *Republic* Plato discusses function, both as exclusive (what something can do alone) and as optimal (what something can do best), in relation to requisite ability for the function. This importantly extends the inquiry into a thing's function since in our efforts to know the function of something we may have to consider what enables that thing to do what it does either exclusively or optimally. Linked with this inquiry into ability and function is virtue as what allows for appropriate use of abilities to perform the function well. Virtue seems to derive from appropriate development of the needed abilities for the function.

Socrates brings up the relation of ability to function in Book II in his creation of the simple city. After he recommends that it might be better for each person to concentrate on doing one job and to thus contribute to the common good he tells Adeimantus,

“It would not, by Zeus, be at all strange (*atopon*),” said I; “for now that you have mentioned it, it occurs to me myself that, to begin with, our several natures are not all alike but different. One man is naturally fitted for one task, and another for another (*phuetai hekastos ou panu homoios hekastô(i), alla diaferôn tèn phusin, allos ep allou ergou praxin*). Don't you think so?” (370a7–b3)

This is the primary principle for assigning different persons and classes to different tasks in the city and presenting analogously the soul with functional parts each performing a different function. Socrates suggests that different people have different natural abilities suited for different functions; human nature seems providentially to divide abilities corresponding to the needs of the arts. But how exactly should we understand natural ability for a task and what are some of the relevant abilities for different tasks?

To begin to answer this question we turn to Plato's preliminary discussion of the natural abilities that make one a suitable guardian in the Kallipolis (374e–375b). He says that people with the appropriate nature to be guards must have the following abilities that we can group into three categories:

- (1) Intellectual: quickly perceptive and pursuing.
- (2) Character: brave due to spiritedness.
- (3) Bodily: strong and swift in body for battle.

We see these rapidly recapped here: “The love of wisdom, then, and high spirit and quickness and strength will be combined for us in the nature of him who is to be a good and true guardian of the state” (376c4–6). Love of wisdom is the intellectual aspect of ability, spirit the character aspect, and speed along with strength, the bodily aspect. In line with the discussion of the tripartite soul, where it is better to have the rational part rule the other parts, Socrates suggests that the intellectual component in the guard's nature balances the character and bodily components making the guard both harsh and gentle and thus suitable for the task.

This classification of natural abilities in a tripartite scheme begins a pattern that we encounter several more times in the rest of the dialogue. When, for example, Socrates selects the rulers of the city and determines whether women are suited for equal education, he similarly locates three sorts of ability that enter into suitability for a function.

In his discussion of the requisite abilities for ruling, Socrates tells us that the rulers, picked out as the best among the soldiers, must be intelligent, capable, and devoted to the city:

“And in this case, since we want them to be the best of the guardians, must they not be the best guardians, the most regardful of the state?” “Yes.” “They must then to begin with be intelligent (*phronimous*) in such matters and capable (*dunatous*), and furthermore careful of the interests of the state (*kêdemonas*)?” “That is so.” “But one would be most likely to be careful of that which he loved (*philôn*).” (412c–d)

Notice the three sorts of abilities that enter the rulers' overall capacity for their task. Socrates hardly elucidates, but clearly intelligence is an expected requisite ability. But the rulers must also be capable or powerful and loving of the city. While capability and power might mean virtuous in character and high-spirited, i.e., strength of character, it could also refer to apparent physical strength and such physical capacities as a compelling voice, while devotion to the city reflects character and spiritedness, since the spirited love what is familiar and are only aggressive to what is unfamiliar and foreign. Thus this underspecified three-pronged qualification may amount to: suitable intellect, good character, and good physical attributes, so that they can see what is best, get others to go along with them, and be motivated to hold to it. The three characteristics for talent or ability at ruling may be compared to the qualifications for talent at a task mentioned in Book II (see 374e–375b) and again in Book V (see 455b and our discussion of this below). Socrates might not clearly mention the body with regard to abilities either because ruling does not so much involve body—neither will philosophizing—or because he can assume that the rulers as best among the soldiers have a suitable body.

Now let us turn to Socrates' consideration of the abilities and function of women in Book V. It is a troubling question regarding women that can be supposed to bother Socrates' interlocutors as much as us. Perhaps it is a good question to test our own ability to challenge our prejudices and spiritedness, and therefore good philosophical preparation. At issue is whether women can perform all the functions (*erga*) that men do, at least so far as the affairs of the city are concerned. To answer this they must consider whether women have the abilities to perform the functions. Abilities to perform a function when well developed are excellences, but prior to this are just abilities requisite for performing the function at all. It seems, then, that function and abilities for the function are coimplicatory, one can only have the function if the abilities are present and where the abilities are present there is the function. Though Socrates does not talk here so much of abilities as of nature, we may see that natural abilities are what is at issue. The relation of nature (or being well-natured, *euphuê*) and abilities seems to be this. If we assume that abilities are somehow innate (though of course to some extent we also have to develop abilities), then nature is the ground



of abilities. The discussion of the nature of women is thus the discussion of ability to have or to perform a function.

Regarding what it means to be naturally capable or well endowed by nature (*euphuê*) Socrates says,

Come then, we shall say to him, answer our question. Was this the basis of your distinction between the one naturally gifted for anything (*euphuê*) and the one not so gifted—that the one learned easily, the other with difficulty; that the one with slight instruction could discover much for himself in the matter studied, but the other, after much instruction and drill, could not even remember (*sôzoito*) what he had learned; and that the bodily faculties of the one adequately served his mind (*dianoia*), while, of the other, the body was a hindrance? Were there any other points than these by which you distinguish the well endowed in every subject and the poorly endowed? (455b–c; Shorey transl. slightly modified.)

Intriguingly, they are trying to deal with the *classes* of women and men, but they look at what seems to endow an individual with relative abilities. Apparently abilities for individuals and kinds are not very distinct and have the same basis. Comparing good nature with poor, they come up with these for good natural ability for a task:

- (1) ready learning and self-discovery beyond what is taught,
- (2) retention of what is learned,
- (3) a suitable body to subserve the thought (*dianoia*).

This list follows the tripartite scheme we encountered earlier, but it might be faulted for leaving out character (cf. 375a–b). Yet this is perhaps covered in preservation of what is learned. Holding on to what is learned is largely a matter of character inasmuch as spiritedness seems crucial for this in Plato's scheme. We might add that these three features Socrates has elicited that go into a good nature apply analogously to functions of all sorts, since for things that do not think, such as inanimate tools, plants, and dumb animals, we still emphasize ready application, retention of capacity for application, and body or structure suitable for the function. We might also notice that the lists of traits constituting ability generally start with intellectual capacity that fits Socrates' insistence that it is better for the individual when the rational part rules the soul. Appropriately, desires, affectivity, should follow cognition. Now let us turn to Socrates' discussion concerning the abilities for and the function of philosophers.

In Book V the claim that cities will not be free from troubles unless philosophers rule or rulers become philosophers (473c–e) results in Socrates explaining just what philosophers are in order to defend his assertion. This is the most elaborate treatment of function and ability in the dialogue. The discussion of what a philosopher is deals with the two parts of the Greek term “philosophy,” namely, “love” and “wisdom.” Socrates distinguishes the philosopher from other people by love for wisdom. But when he defines the philosopher, is he defining a function, philosophizing or loving wisdom, or the capacity to perform such a function? If a function is being determined, then Socrates has to say what the philosopher does best or alone, whereas if this is a capacity to perform a function this resembles those other

places where he figured out what constitutes ability for a function. Perhaps because what he defines is “philosopher,” i.e., *lover* of wisdom, he ends up discussing both the function and the requisite abilities for it. Socrates views the pursuit of wisdom as the philosopher’s function, and he clarifies the abilities allowing for this.

Somewhat surprisingly Socrates begins with the affective side: he starts defining the philosopher by noting that when we call someone a lover of X this does not mean that he loves some of X but not some other part of X, but he loves all of X (474c). Thus the philosopher loves all of wisdom or all instances of it, not merely some part of it. So far he seems to be defining philosophy as a function: pursuit of all wisdom. But then he emphasizes how intensely the lover loves and how insatiable he is (see 475c), so that “loving all of wisdom” shifts over to a huge, intense love. This appears, therefore, to be an account of what gives the ability to be a philosopher as well as an account of the function. Then Socrates adds that the philosopher, unlike lovers of other things such as sights or crafts, loves knowing what is, what is always the same, i.e., the Forms (476a ff.). This specifies the content of the wisdom, and hence the wisdom pursued, clarifying the philosopher’s function.<sup>8</sup>

In the beginning of Book VI, Socrates is much more explicit and expansive regarding the requisite abilities for philosophy as the function of loving all of wisdom about the Forms. Socrates builds to an analysis of the abilities for the function of the philosopher when he says that

“Then, as we were saying at the beginning of this discussion, the first thing to understand is the nature that they must have from birth; and I think that if we sufficiently agree on this we shall also agree that the combination of qualities that we seek belongs to the same persons, and that we need no others for guardians of states than these.” “How so?” “We must accept as agreed this trait of the philosophical nature (*tôn philosophôn phuseôn*), that it is ever enamored of the kind of knowledge which reveals to them something of that essence which is eternal (*tês ousias aei ousês*), and is not wandering between the two poles of generation and decay.” “Let us take that as agreed.” “And, further,” said I, “that their desire is for the whole of it (*pasês autês*) and that they do not willingly renounce (*hekotes aphientai*) a small or a great, a more precious or a less honored, part of it.” (485a–b)

The abilities of the philosopher connect with the fervent love for what always is. Reference to their devotion to the whole or all of learning brings in the consideration of the ability for philosophy. In a remarkable extended discussion that provokes Adeimantus’ complaint about Socrates’ mode of argumentation (see 487a–d), Socrates purports to show that philosophers have all the moral virtues and an array of abilities. It appears that he is deriving from the account of the philosopher’s function that the philosopher has all the conventional moral virtues and the sorts of abilities that are requisite for any function and the function of philosophy in particular. The passage runs:

“Consider, then, next whether those who are to meet our requirements must not have this further quality in their natures.” “What quality?” “The spirit of truthfulness (*apseudeian*), reluctance to admit falsehood in any form, the hatred of it and the love of truth.” “It is likely,” he said. “It is not only likely, my friend, but there is every necessity that he who is by nature enamored of anything should cherish all that is akin and pertaining to the object of his love.” “Right,” he said. “Could you find anything more akin to wisdom than truth?”

“Impossible,” he said. “Then can the same nature be a lover of wisdom and of falsehood?” “By no means.” “Then the true lover of knowledge must, from childhood up, be most of all a striver after truth in every form.” “By all means.” “But, again, we surely are aware that when in someone the desires incline strongly to any one thing, they are weakened for other things. It is as if the stream has been diverted into another channel.” “Surely.” “So, when someone’s desires have been taught to flow in the channel of learning and all that sort of thing, they will be concerned, I presume, with the pleasures of the soul in itself, and will be indifferent to those of which the body is the instrument, if someone is a true and not a sham philosopher.” “That is quite necessary.” “Such a one will be temperate (*sôphrôn*) and by no means greedy for wealth; for the things for the sake of which money and great expenditure are eagerly sought others may take seriously, but not he.” “It is so.” “And there is this further point to be considered in distinguishing the philosophical from the unphilosophical nature.” “What point?” “You must not overlook any touch of illiberality (*metechousa aneleutherias*). For nothing can be more contrary than such pettiness (*smikrologia*) to the quality of a soul that is ever to seek integrity and wholeness in all things human and divine.” “Most true,” he said. “Do you think that a mind habituated to thoughts of grandeur (*megaloprepeia*) and the contemplation of all time and all existence can deem this human life a thing of great concern?” “Impossible,” said he. “Hence such a one will not suppose death to be terrible?” “Least of all.” “Then a cowardly and illiberal spirit, it seems, could have no part in genuine philosophy.” “I think not.” “What then? Could a person of orderly spirit, not a lover of money, not illiberal, nor a braggart nor a coward, ever prove unjust, or a driver of hard bargains?” “Impossible.” “This too, then, is a point that in your discrimination of the philosophic and unphilosophic soul you will observe—whether the person is from youth up just and gentle (*dikaia te kai hêmeros*) or unsocial and savage.” “Assuredly.” (485b–486b; Shorey transl. slightly modified)

We thus see how from the love of wisdom and truth Socrates has seemed to derive all of the moral virtues for the philosopher. Starting off talking about love of truth in order to gain the principle that love extends to what is akin to what is loved and that what the philosopher loves must be genuine rather than false, he finds all the virtues of character in the philosopher. These virtues presuppose some success at performing the function of philosophizing. Due to the philosopher’s devotion to truth, he has little time or interest in more unreal things. We may observe, however, that these virtues more resemble the conventional accounts of the virtues than the accounts developed previously in Book IV, and this may contribute to the provocation of Adeimantus. Genuine love of wisdom and truth involves a certain orientation of character on the part of those who adequately fulfill the function. Even if genuine philosophers might also have the true cardinal virtues, the function presupposes and accords with the kind of orientation of character that Socrates reveals: devotion to the real things rather than mere imitations of them. And this adequately links the conventional virtues of character with the philosopher.

And now Socrates finally goes on to the abilities definitely required for the philosopher:

“Nor will you overlook this, I fancy.” “What?” “Whether he is quick or slow to learn (*eumathês ê dusmathês*). Or do you suppose that anyone could properly love a task which he performed painfully and with little result from much toil?” “That could not be.” “And if he could not keep (*sôzein*) what he learned, being steeped in oblivion, could he fail to be void of knowledge?” “How could he?” “And so, having all his labor for naught, will he not finally be constrained to loathe himself and that occupation?” “Of course.” “The forgetful soul, then, we must not list in the roll of competent lovers of wisdom, but we require a good

memory (*mnēmōnikēn*).” “By all means.” “But assuredly we should not say that the want of harmony and seemliness in a nature conduces to anything else than the want of measure and proportion.” “Certainly.” “And do you think that truth is akin to measure and proportion or to disproportion?” “To proportion.” “Then in addition to our other requirements we look for a mind endowed with measure and grace (*emmetron kai eucharin*), whose native disposition will make it easily guided (*euagōgon*) to the aspect of the ideal reality in all things.” “Assuredly.” “Tell me then, is there any flaw in the argument? Have we not proved the qualities enumerated to be necessary and compatible with one another (*hepomena allēlois*) for the soul that is to have a sufficient and perfect apprehension of reality?” “Nay, most necessary,” he said. “Is there any fault, then, that you can find with a pursuit which someone could not properly practice unless he were by nature of good memory, quick apprehension, magnificent, gracious, friendly and akin to truth, justice, bravery and sobriety (*phusei eîē mnēmōn, eumathēs, megaloprepēs, eucharēs, philos te kai suggenēs alētheias, dikaiosunēs, andreias, sōphrosunēs*)?” (486b–487a)

Socrates has philosophy imply the intellectual ability of quick learning. The good retention of what is learned might be an intellectual ability or akin to the sort of spiritedness that keeps one holding strongly to something. We see how these two sorts of ability fit well with what Socrates has said previously about the kinds of requisite abilities. But since he is speaking of requisite abilities of the *soul*, he cannot simply speak of the body.

The measure and grace of the soul seems to be what steers it toward the Forms, but it also soon becomes the conventional moral virtues. It turns out, then, that this surprisingly resembles the good condition of the body because it provides good appearance and a kind of fitness for the function. Thus, loving all wisdom as the true philosopher does (where this wisdom is involved in truth) is a *sufficient condition* for having the requisite abilities, and consequently such abilities are the *necessary* conditions for truly loving wisdom. Surely quick learning and retention recall to us the earlier lists of abilities. And either philosophy as Socrates conceives it places more definite demands upon character than upon the body, and so he substitutes grace, harmony, and the moral virtues for a fit body, or the conventional moral virtues really amount to a suitable control of the body for undertaking philosophy. Hence we may say that the extended derivation of conventional sorts of virtues from the genuine philosophical function turns out to be in fact the discovery of the necessary condition for enough bodily control to philosophize. Hopefully when real philosophy is achieved, such ability turns into the true cardinal virtues.

Yet despite all the talk of the virtues and abilities of the philosopher, Adeimantus objects to the mode of argument used by Socrates, and he complains that in reality philosophers become weird (*allokotous gignomenous*), all wicked (*pamponērous*), or useless (*achrēstous*, 487d). This leads Socrates in meeting these charges to argue through famous images that those genuinely having these abilities and capable of performing the function should hardly be useless, though the community may not properly utilize them, as in the ship of state image the person who can truly navigate is ignored. And the fine abilities can be turned toward disreputable purposes rather than to good purposes, even toward tyranny instead of philosophy, when the conventional aims of the crowd or intimates turn the gifted youth away

from his true function. This confirms the significance and pattern of the reflection upon abilities in relation to function.

### 12.3 Ability and Desire

In addition to presenting the requisite abilities for philosophy, we see that in setting out the abilities needed for philosophy Socrates questions whether someone could persistently desire to pursue wisdom without the ability to make any progress in the enterprise. Socrates points out that if someone learns with difficulty (*dusmathês*) and has bad memory (*lêthês hôn pleôs*), then such a person will not have lasting desire for wisdom (486c). Thus suitable abilities are a necessary condition for having lasting desire. Loving all wisdom requires the abilities to pursue it, i.e., the necessary conditions for it. It therefore seems that desires (such as love of wisdom) depend upon abilities (cf. 487a). At least ability is a necessary condition for *lasting* rather than ephemeral desire since we generally give up our desire if its realization is hopeless and we recognize this.<sup>9</sup>

We find some confirmation in Book VII where Socrates discusses the education of philosophers and in the middle of repeating some of the requisite abilities for ruling and philosophy he says,

“And we must demand a good memory and doggedness (*arraton*) and industry in every sense of the word (*pantê(i) philoponon*). Otherwise how do you suppose anyone will consent both to undergo all the toils of the body and to complete so great a course of study and discipline?” “No one could,” he said, “unless most happily endowed.” “Our present mistake,” said I, “and the disesteem that has in consequence fallen upon philosophy are, as I said before, caused by the unfitness of her associates and wooers. They should not have been bastards but true scions.” (535c)

Socrates points out that in order for someone to be willing to endure the labor involved with learning and to make much progress he or she needs to have the ability for such a task, i.e., be naturally suited for it. Thus, once again, lasting desire for philosophizing or anything or any function depends on having the requisite ability for it. Socrates does not mean to suggest that all it takes for lasting desire for possessing something or engaging in some function is ability for it, but rather that such ability is necessary (not sufficient) for lasting desire. Other conditions enter into making desire for something lasting. In the case of philosophy for example, the pursuit of it may further depend on education and the type of political regime and whether this allows for its pursuit.<sup>10</sup>

Some of the deep significance of this fairly obvious point that lasting desire to perform some function depends upon ability to do it may now be illustrated. We recall Glaucon’s usage of the ring of invisibility seemingly to prove that we all harbor secret natural desires to outdo others, i.e., we are all naturally unjust. Glaucon supposed that the ring of invisibility discloses our repressed desires for *pleonexia*, repressed because in confrontation with others we have to prefer our own continued existence and act justly unwillingly. Once we get the ring of invisibility that provides us the ability to satisfy our desires, however, our true desires

can emerge. Glaucon suggests that desires are driving our lives, but Socrates' view that desires depend upon abilities indicates we cannot truly love something without the ability for it and without practicing it. So in fact abilities drive our lives in conjunction with desires that accord with them. Glaucon's tale, then, by changing ability through providing the ring of invisibility, is not exposing our true desires, our heart of hearts, but changing the recipient's loves. With new abilities we develop new desires; hence the ring story is not actually revealing what is natural at all. The magical ring that allows for ability that humans do not have naturally thereby gives rise to unnatural desire, the desire to unjustly outdo others. But since abilities do or should drive human life, Socrates has justice centering on abilities rather than desires.

Now of course ability may not lead to love: the gifted in something may hate it. But there cannot be any love without ability, ability being a necessary condition for sustained desire. Hence it is at least in principle possible that persons with ability to do a task can be led to desire to do the task. So in this sense the city of the *Republic* is hardly impossible, if we assume the requisite abilities are available. In the case of the genuine philosophers their natural abilities and natural desires conform perfectly to each other. Socrates offers the optimistic view that those who do not have the requisite ability will be discouraged from the task. When he later speaks of those who rush to philosophy not being qualified, we doubt that they really love all wisdom that is about the truth.

We may observe in modern political thought a view rather like that in Glaucon's speech. For Hobbes all humans, or at least some, are taken to harbor such desires to gain advantage over others that there is a war of all against all. Even those not so inclined therefore have to protect themselves against those who are. Subsequent thinkers, such as Locke and Rousseau, return to the Platonic insight about ability directing desire to avoid some of the Hobbesian picture, but now only through the new emphasis upon historical development. The most primitive humans or those in simple societies do not have great desires because they have no ability to satisfy them; through time and the development of the arts and the apparatus needed for them, however, human abilities so expand that their desires may also expand commensurately.<sup>11</sup> These philosophers seem to escape attacking human nature as in Hobbes through seeing the expansion of human desires to be expected with the historical expansion of human abilities.<sup>12</sup> The passages that preface this paper show how common it is for modern philosophers to resist the very basis of the scheme in the *Republic*, that is, they resist having humans differ in abilities and hence to have different functions. Instead abilities are assumed to be basically the same or desires are assumed to give rise to abilities rather than the reverse. Instead of having power and ability intrinsic to natural beings, the modern tendency is to have power and ability develop from external relations.

We hope that we have supported Jerry Santas' emphasis upon Plato's use of function in the *Republic*. We have tried to show that function plays even a larger role than Jerry suggests and that further consideration of its linkage with abilities, virtue, and desire offers considerable light upon Plato's project in the dialogue and has much intrinsic interest.

## Notes

1. By understanding that the Forms do have causal functions we can possibly explain how what Santas calls “the theory of the form of the good” is not an isolated theory of goodness but part of the functional theory of goodness. Santas holds that “One of the most puzzling features of the *Republic* is that Plato does not apply the theory of the form of the good to determine what is a good man and a good city. So the theory of the form of the good sits there, in the middle of the *Republic*, idle” (p. 214). Cooper (1997, p. 23) observes that the form of the good has functions, or as he calls them “functional properties,” such as being the cause of the goodness of other things.
2. Kraut (2005) in his review of Santas’ *Goodness and Justice: Plato, Aristotle, and the Moderns* argues, contra Santas, that Plato need not use any of the parts of the functional theory of goodness from *Republic* I in the rest of the dialogue. Kraut insists that it is a mistake to suppose that the functional theory of good is used in Plato’s account of the good city, the virtues, and the argument in Book IV that justice is superior to injustice (p. 453). To support this claim Kraut argues that:

- (1) The new beginning in the discussion in Book II suggests that we should not suppose that any of the premises from Book I are accepted by any of the interlocutors from Book II onward simply because they were accepted in Book I. We should think rather that the characters of the dialogue reject the arguments in defense of justice in Book I and insist on new arguments in its defense from Book II onward (p. 453).
- (2) In drawing the conclusion in Book IV that it is far better to be just rather than unjust, Socrates does not rely on the assumption that living is a function of the soul. “So one of the essential premises of the ‘function’ argument of I has been dropped” (p. 452).
- (3) No passages from Book IV in the Greek have *ergon* when referring to the parts of the soul but the words “function” and “work” are only supplied by translators.
- (4) None of the premises from the defense of justice in Book IV relies on the functional theory of the good as presented in Book I since Plato “defends the one-man-one-job principle on the grounds that when each craftsman or farmer sticks to his job, they will be able to produce more and better goods for others” (p. 453). Moreover, when Socrates reaches the conclusion that the rational part of the soul should rule “he does so on the grounds that this part of the soul is wise and exercises foresight on behalf of the whole soul” (p. 453). Kraut continues that there is no textual evidence to support the idea that reason should rule because it is its function to do so, but rather it should rule only because it is wise.

Kraut’s (1) is unconvincing since what Glaucon rejects in the beginning of Book II are the arguments in defense of justice, not all the premises in the arguments. And (2) is unpersuasive since Socrates in arguing that justice is superior to injustice in Book IV drops only one of the functions of the soul he mentions in Book I but he hardly drops functional talk. Point (3) is insufficient to show that function drops out of the picture and even Kraut admits this (pp. 452–3). Finally, (4) overlooks how the whole division of tasks for parts of the soul and classes of the city is based on different functions and abilities. Cooper, in Kraut (1997), helpfully contends that the function of reason is to rule *with wisdom* which suggests that reason does have the function of ruling since it has requisite ability or virtue for such a task, i.e., wisdom (p. 19).

3. Socrates uses the notion of function as early as 335d ff. in his fourth refutation of Polemarchus’ account of justice where he attempts to show that it cannot be the role of a just man to harm anyone. He argues that if when harming a horse or a dog we must do so with regard to their excellence or virtue, then we must harm their body or soul in such a way that it interferes with their doing their function according to virtue. We should notice that Socrates holds that it is not the function of justice to harm and that all sorts of things have functions that we can



determine; he mentions the function of heat (heating), music (to make one musical), justice and goodness (to make one good or just), and animals.

4. Santas (2001) calls these helpfully “exclusive” and “optimal” functions (p. 69). Function  $x$  is the exclusive function of  $y$  if only  $y$  can do it. Function  $x$  is the optimal function of  $y$  if  $y$  can do it best among the things which can do  $x$ . Both Plato and Aristotle use these senses of function. For Aristotle's discussion of function peculiar to humans in an ethical context, see *NE* 1097b22–1098a18. Santas points out that in the *NE* Aristotle uses only Plato's notion of exclusive function when he claims that humans must have a distinctive function, i.e., a function which only they can do. Aristotle uses Plato's notion of optimal function in his consideration of who should rule (p. 237). We should add to this that in the *NE* Aristotle determines human function not so much by what we uniquely do but by what enables us to do all that is uniquely human. All characteristically human action derives from speech or thought. Santas hints at this on p. 244.
5. Santas (2001) seems to have this wrong on p. 72 and then pp. 80–1 where he supposes that optimal functioning could be the best that the individual can do among the things it can do, in which case the truly untalented at much of anything end up with a function.
6. On the connection of function to happiness in the *Republic*, see also Irwin (1995, p. 253). He claims that Plato connects function with virtue explicitly and function with happiness implicitly. Irwin also points out that there are different ways to understand function, that is, as a task something was designed to perform, as a task something is expected to perform, or as Plato uses it in the *Republic* in relation to humans, as something a thing does naturally and necessarily given its nature (p. 252). Using necessity here seems misleading insofar as not all functions that belong to something by nature will be performed.
7. The non-eudaemonistic life or alternative eudaemonistic approaches, such as hedonism and desire satisfaction, will not so successfully integrate function into determination of the good.
8. The love involved in the philosopher could be clarified in ways familiar to everyone because they could speak of various comparable lovers of boys, of wine, of food, of honor. But now they cannot so easily understand wisdom or the truth by comparing it to sights and sounds. So this particular part of the argument aims at people such as Glaucon rather than just anyone (or even the lovers of sights and sounds). Socrates addresses himself to Glaucon (475e6–7) and finally introduces the theory of Forms explicitly. Perhaps the clarification of the philosopher's function demands that one have the abilities of the philosopher to be ready for the function, i.e., to be best at it or alone to have it. And the limiting of the objects of wisdom to non-perceptible intelligible objects suggests the intellectual abilities requisite for philosophizing.
9. But see 497e on wishes, and consider our impossible wishes such as that the past could be different, e.g., our desire to be together with a lost, deceased love. Perhaps such wishes hardly, however, lead to much action unless they connect to a project of possible achievement.
10. On the role of political constitutions and education in the development or frustration of natural abilities in relation to desire, see Lear (1997). Lear points out that the soul internalizes cultural values, attitudes, and desires, while at the same time externalizing values, attitudes, and desires into the political community it finds itself in. This reciprocal process of internalization and externalization between the soul and the city clearly affects whether natural abilities are developed or frustrated and the sorts of desires citizens may have.
11. See Rousseau's critique of Hobbes and others in the *Discourse on the Origins of Inequality* (1987, p. 38), where he claims that they take the developed characters of humans in society and project them mistakenly in their descriptions of humans in the state of nature.
12. See Locke in the *Second Treatise of Government*, sections 36–37, regarding the situation before money is introduced, i.e., before there is money and ability to acquire, there will be no real desire for large acquisition. While Plato is speaking of the psychology of the individual, Locke is on the trail of historicity and how social conditions open up the possibility of changes in human desires and even human “nature.” Rousseau's *Discourse on the Origin of Inequality* traces the human development of pride along with expanding of civilization and technical abilities.

## Bibliography

- Cooper, J.M. 1997. "The Psychology of Justice in Plato." In edited by R. Kraut, 17–30.
- Hobbes, T. 1974. *Leviathan*. London: Collier Books.
- Irwin, T. 1995. *Plato's Ethics*. New York: Oxford University Press.
- Kraut, R., ed. 1997. *Plato's Republic: Critical Essays*. Oxford: Rowman and Littlefield.
- Kraut, R. 2005. Review of Santas' *Goodness and Justice: Plato, Aristotle, and the Moderns*. *Ancient Philosophy* 25:446–63.
- Lear, J. 1997. "Inside and Outside the Republic." In edited by R. Kraut, 61–94.
- Locke, J. 1980. In *Second Treatise on Government*, edited by C.B. Macpherson. Indianapolis: Hackett.
- Rousseau, J.-J. 1987. *Basic Political Writings of Jean-Jacques Rousseau*. Indianapolis: Hackett.
- Santas, G. 2001. *Goodness and Justice: Plato, Aristotle, and the Moderns*. Oxford: Blackwell Publishers.
- Sartre, J.-P. 1987. "Existentialism." In *Existentialism and Human Emotions*, edited by J.-P. Sartre. New York: Citadel Press.

# Chapter 13

## Knowledge, Virtue, and Method in *Republic* 471c–502c

Hugh H. Benson

The last third of Plato's *Meno* has been an enigma for readers of Plato for centuries. After leading Meno to the recognition of his own ignorance (of the nature of virtue) and persuading Meno that rectifying that ignorance is possible (by offering the theory of recollection and the conversation with the slave in response to Meno's paradox), Socrates sets about determining whether virtue is teachable. He first offers an argument to the effect that it is teachable on the grounds that virtue is knowledge and knowledge is teachable. Next he offers an argument that it is not teachable, on the grounds that what is teachable has teachers and there are no teachers of virtue. He concludes by indicating that the first argument contains a flaw and maintaining that virtue is true belief acquired by divine dispensation.

What readers have found so puzzling about this argument in the last third of the *Meno*<sup>1</sup> is not so much that Socrates appears here to be arguing, on the one hand, for *p* and then, on the other, for *not-p*.<sup>2</sup> Rather, what is puzzling is that Socrates appears to plump for the second argument. This is puzzling for at least three reasons. First, in plumping for the second argument, Socrates appears to be abandoning his intellectualism—the view minimally that knowledge (of the good) is necessary and sufficient for virtue—maintained repeatedly in the other Socratic dialogues.<sup>3</sup> True belief according to Socrates in the last third of *Meno* appears sufficient for virtue.<sup>4</sup> Second, Socrates had just asked to employ an alleged mathematical method in order to seek an answer to Meno's question whether virtue is teachable—the so-called method of hypothesis. Having received Meno's agreement to do so, he employs that method in the first argument to the effect that virtue is knowledge and so teachable. But then he appears to abandon the method he had just been given permission to employ in the second argument to the effect that virtue is not teachable and so not knowledge. Third, this last argument has appeared to many to be subject to a variety of fallacies, perhaps the most crude of which is an equivocation between teachable and taught.<sup>5</sup>

---

H.H. Benson (✉)  
University of Oklahoma, Norman, OK, USA  
e-mail: hbenson@ou.edu

These difficulties with the second argument have led a variety of commentators to maintain that Socrates does not take this second argument seriously.<sup>6</sup> But I have argued elsewhere that we have good reasons to doubt that Socrates abandons his method of hypothesis in this second argument. Rather this second argument conforms to one of two processes described at *Phaedo* 101d1–e3 for confirming a hypothesis whose consequences answer the question at hand—roughly the process of determining whether the consequences of the hypothesis are consistent with the facts on the ground and other background beliefs or information concerning the topic under discussion—the argument *from* the hypothesis.<sup>7</sup> If this is right, then we cannot so easily dismiss the second argument in the *Meno*. It may have an un-Socratic conclusion and it may even be subject to a variety of fallacies, but it is presented as part of a general philosophical method that Socrates is made to recommend.<sup>8</sup> What then are to make of this argument in the *Meno*?

I want to throw light on this question by turning to a rather surprising text—*Republic* V–VI from the beginning of the so-called third wave at 471c through the introduction of the Form of the Good at 502c.<sup>9</sup> I will maintain that the content and structure of the argument during this stretch of the *Republic* parallels the last third of the *Meno*. The similarities highlight a difference and the difference suggests an explanation of what is unsatisfactory in the *Meno*. My focus in this essay, however, is not to claim that the passage in the *Republic* determines or even informs how we should understand Plato's views about virtue, knowledge, or its teachability in the *Meno*. For the purposes of this essay, I leave these issues wholly to one side. Rather the primary focus of this essay is the light the *Republic* passage throws on the philosophical method recommended and pursued in the *Meno*. What is unsatisfactory, if anything, about the application of the method pursued in the *Meno*?

I will argue that in response to the question whether Kallipolis is possible Socrates employs the method of hypothesis recommended in the *Meno*, the *Phaedo*, and the *Republic*.<sup>10</sup> Just as in the *Meno* where the question whether virtue is teachable is reduced to the question whether virtue is knowledge, so in the *Republic* the question whether Kallipolis is possible is reduced to the question whether philosophy and political power coincide (*Rep.* 473b–e). Next Socrates sets out to confirm the truth of the hypothesis that philosophy and political power coincide. First, he provides an argument *to* the hypothesis—going up to a higher hypothesis that philosophers are knowers of Forms and then arguing that so understood philosophers have the necessary and sufficient characteristics for ruling (*Rep.* 474c–487a). Next, he provides an argument *from* the hypothesis—testing the claim that philosophy (so understood) and political power coincide against experience, *endoxa*, and other ordinary beliefs (*Rep.* 487b–502c).<sup>11</sup> At this point, however, the parallel between this passage and the last third of the *Meno* breaks down. In the *Meno*, the hypothesis that virtue is knowledge is discredited by the argument *from* the hypothesis. As Socrates puts it, a consequence of the hypothesis is that there are teachers of virtue, but in fact there are no such things. In the *Republic*, however, the hypothesis that philosophy and political power coincide is further confirmed by the argument *from* the hypothesis. Seeing how the hypothesis in the *Republic* escapes refutation in the argument *from* the hypothesis will illuminate the precise nature of this procedure

and clarify its roughly *a priori* character. In being confirmed by both procedures in the *Republic*, unlike the *Meno*, progress along the longer road to the unhypothetical first principle of everything has been made.<sup>12</sup>

The essay falls into four parts: First, I will briefly describe the method of hypothesis as it is articulated in the *Meno*, the *Phaedo*, and the *Republic* (Section 13.1). Next, I will equally briefly outline the structure of the method as it is practiced in the last third of *Meno* (Section 13.2). In these two sections I will be rehearsing and elaborating on previous work.<sup>13</sup> Third, I will present in some detail the argument structure of *Republic* 471c–502c. I will highlight both various allusions to the last third of the *Meno*, and more importantly its structural parallel with the method practiced in this portion of the *Meno*. I will also underscore the salient structural difference from the last third of the *Meno* (Section 13.3). Finally, I will speculate on the significance of this difference between these arguments of the *Republic* and the *Meno* as it pertains to Plato's recommended method of philosophical inquiry (Section 13.4). I will leave it to others and another time to speculate about the significance of this difference as it pertains to Plato's views about virtue, knowledge and its teachability in the two dialogues.

### 13.1 The Method of Hypothesis

As I mentioned above, after Socrates has persuaded Meno of his ignorance (by means of the *elenchos* in the first third of the dialogue) and the possibility of rectifying that ignorance despite their mutual ignorance (by means of the theory of recollection and an example of a geometrical conversation with a slave in the middle third of the dialogue), he sets out somewhat reluctantly to determine whether virtue is teachable (in the final third of the dialogue). His reluctance stems in part from his preference to pursue first what virtue is, but he accedes to Meno's insistence to return to the question with which the dialogue began on the condition that Meno permit him to employ a method borrowed from the mathematicians.<sup>14</sup> Here for the first time in the dialogues Socrates refers rather explicitly to what has come to be known in the literature as the method of hypothesis.<sup>15</sup> In light of the geometrical example Socrates uses to introduce the method—the details of which are obscure, controversial, and do not need to be pursued here—the method appears to consist of the following two stages. The first stage is to identify a hypothesis such that its truth is necessary and sufficient for a determinate answer to the question under consideration. It consists in identifying a hypothesis from which the answer to the question under consideration can be derived.<sup>16</sup> In the case of the geometrical example, the hypothesis appears to be that the area is “such that when one has applied it as a rectangle to the given straight line in the circle it is deficient by a figure similar to the very figure which is applied,” while in the case of the question whether virtue is teachable the hypothesis is that virtue is a kind of knowledge (cf. *Meno* 87b5–c7). The second stage is to confirm the hypothesis or determine whether the hypothesis is true. One seeks to determine whether the given area is “such that when one has applied it as a rectangle to the given straight line in the circle it is deficient by

a figure similar to the very figure which is applied” or whether virtue is a kind of knowledge.<sup>17</sup>

Socrates does not tell us much in the *Meno* about how these stages are to be performed.<sup>18</sup> In the *Phaedo*, however, he does provide some substance to the confirmation stage.<sup>19</sup> In reply to Cebes’ objection to Socrates’ third attempt to prove the immortality of the soul, Socrates explains that he must engage in “a thorough investigation of the cause of generation and destruction” (*Phd.* 95e9–96a1; Grube transl.). After describing some failed methods for pursuing this investigation, Socrates returns to the method of hypothesis, saying:

I started in this manner: taking as my hypothesis (*hupothemenos*) in each case the theory that seemed to me the most compelling (*errômenestaton*), I would consider as true, about cause and everything else, whatever agreed (*sumphônêin*) with this, and as untrue whatever did not so agree. (*Phd.* 100a3–100a7; Grube transl.)

Here Socrates describes the first stage mentioned in the *Meno*—the reduction stage or the stage of identifying a hypothesis from which an answer to the question under consideration can be derived. A bit later he describes the confirmation stage or the stage of determining whether the hypothesis itself is true as follows:

You would ignore [an individual who questioned the hypothesis] and would not answer until you had examined whether the consequences that follow from it agree with one another or contradict one another. And when you must give an account of your hypothesis itself you will proceed in the same way: you will assume another hypothesis, the one which seems to you best of the higher ones until you come to something acceptable, but you will not jumble the two as the debaters do by discussing the hypothesis and its consequences at the same time, if you wish to discover any truth. (*Phd.* 101d3–e3; Grube transl.)

Here Socrates goes beyond anything described in the *Meno* in explaining that the confirmation stage consists of two procedures—which I have called elsewhere “an argument *to* the hypothesis” and “an argument *from* the hypothesis.” The argument *to* the hypothesis consists of “assuming another hypothesis, the one which seems to you best of the higher ones until you come to something acceptable” from which the original hypothesis being confirmed can be deduced. The argument *from* the hypothesis consists of “examining whether the consequences that follow from it agree with one another or contradict one another.” Nevertheless, despite offering more substance, Socrates leaves much unexplained and full of puzzles.

Finally, in the *Republic* Socrates is made to describe the greatest study—the Form of the Good (*Rep.* 505a2). Socrates calls the method by which one seeks to acquire this knowledge—the longer road—dialectic (*dialektikên*) (*Rep.* 532b4), but it is simply an elaboration of the method we have found in the *Meno* and the *Phaedo*. In distinguishing dialectic from *dianoetic*—the method practiced by the mathematicians that corresponds to the mental state of *dianoia* in the third section of the Divided Line—Socrates explains that both dialectic and *dianoetic* employ hypotheses. The difference is that *dianoetic* “uses as images the things that were imitated before” and “proceeds not to a first principle but to a conclusion,” while dialectic proceeds “without the images used in the previous section, using Forms themselves and making its investigation through them” and “makes its way to a first principle that is not a hypothesis.” In light of what we have seen in the *Meno*

and *Phaedo* we might explain this difference as follows: both dianoetic and dialectic seek to answer the question at issue by identifying a hypothesis from which an answer to the question can be derived. Dianoetic, however, limits itself to this first reduction-stage of the hypothetical method—identifying the hypothesis and showing how the hypothesis entails the answer to the question, choosing the hypothesis that is generally consistent with sense experience and common sense.<sup>20</sup> Dialectic, on the other hand does not limit itself to the first stage of identifying and drawing out the consequences of the hypothesis, but continues with an extensive confirmation stage—both employing the procedure of an argument *to* the hypothesis all the way up to the “unhypothetical first principle of everything” and the procedure of an argument *from* the hypothesis, but this time in some way not “employing” sense experience and common sense but making its way using Forms alone. At 534b8–c3 Socrates elucidates this last procedure a bit further saying that one must be able to (1) define the Form of the Good in a *logos* distinguishing it from everything else, (2) go through all the examinations as if in a battle, (3) examine not according to *doxa*, but according to being, and (4) survive all of this with the *logos* undefeated.<sup>21</sup>

Even now much remains unexplained and in need of further elucidation.<sup>22</sup> But we cannot, I think, escape the fact that a rough outline of a method has emerged in these three dialogues. The method by which one seeks knowledge of the answer to a given question—whether virtue can be taught, whether the soul is immortal, or whether justice is a “good we like for its own sake and also for the sake of what comes from it” (357c1–2; Grube/Reeve transl.)—consists of two stages:

- (1) First (a) one seeks to identify a hypothesis from which an answer to the question can be derived and (b) one shows how the truth of the hypothesis entails the answer to the question.
- (2) Next one seeks to confirm the truth of the hypothesis, (a) first by identifying a further hypothesis from which the original hypothesis can be derived and showing how this derivation goes until one reaches the “unhypothetical first principle of everything,” and then (b) by testing the hypothesis in some kind of *a priori* way.<sup>23</sup>

We will see in the next sections that both the argument aimed at determining whether virtue is teachable in the *Meno* and the argument aimed at determining whether Kallipolis is possible in the *Republic* display this rough structure. Consequently both arguments amount to applications of the method of hypothesis and both can supplement our understanding of the method so far described.

### 13.2 The Structure of *Meno* 87b2–96d4

At *Meno* 87b2–c9, Socrates identifies the hypothesis from which an answer to the question whether virtue is teachable can be derived. Socrates maintains that virtue is teachable just in case virtue is a kind of knowledge.<sup>24</sup> The derivation of the answer from the hypothesis that virtue is a kind of knowledge is straightforward. One need



only supply the claim that all and only knowledge is teachable (*Meno* 87c2–6). Consequently, if the hypothesis that virtue is a kind of knowledge is true, then the answer to the question at issue is that virtue is teachable. If the hypothesis is false, then the answer is that virtue is not teachable. The bulk of this portion of the *Meno* is concerned with confirming the truth of the hypothesis—the second stage of the method of hypothesis.

At *Meno* 87d2–89c4 Socrates proceeds to the aspect of this stage that I have labeled “the argument *to* the hypothesis.” Socrates begins by hypothesizing that virtue is good (*Meno* 87d2–3), and then providing an argument to the effect that nothing else is good other than knowledge (*Meno* 87e1–89a2). He concludes from this that virtue is a kind of knowledge (*Meno* 89a3–4). Here we have a fairly explicit example of the aspect of the confirmation process which in the *Phaedo* is described as “assuming another hypothesis [i.e., that virtue is good], the one which seems to you best of the higher ones until you come to something acceptable [which the *Meno* describes as ‘remaining for us’]” from which the original hypothesis [i.e., that virtue is a kind of knowledge] being confirmed can be deduced. The truth of the hypothesis that virtue is a kind of knowledge has been confirmed by means of an argument *to* the hypothesis. But our review of the method of hypothesis in the *Meno*, *Phaedo*, and *Republic* indicates that such confirmation is insufficient.<sup>25</sup> The hypothesis must also be confirmed by means of an argument *from* the hypothesis. And it is to this portion of the confirmation stage that Socrates turns at *Meno* 89c.

When Meno agrees at 89c2–4 that virtuous individuals become virtuous not by nature, but by learning or being taught given the truth of the hypothesis that virtue is knowledge, Socrates expresses some misgivings, which he explains as follows: “I am not saying that it is wrong to say that virtue is teachable if it is knowledge, but look whether it is reasonable of me to doubt whether it is knowledge” (*Meno* 89d3–6; Grube transl.). Socrates here expresses misgivings not about the first stage of the method of hypothesis as it has been performed in the *Meno*. The claim that virtue is teachable just in case it is a kind of knowledge remains undisturbed. What Socrates expresses concern about is whether the claim that virtue is a kind of knowledge has been adequately confirmed by the argument *to* the hypothesis. He is concerned that the confirmation of this hypothesis by the argument *from* the hypothesis will have different results. He is concerned, that is, that the hypothesis that virtue is a kind of knowledge is incompatible with other background beliefs. In particular, Socrates maintains the general principle that something is teachable just in case it has teachers and learners (*Meno* 89d6–e3), and then expresses his inability to discover teachers of virtue (*Meno* 89e6–9). His inability to discover such teachers is reinforced by a series of arguments first with Anytus and then with Meno undermining the authenticity of the claim of traditional and alleged teachers of virtue—the Sophists, the *kaloï kagathoi*, and the poets. What is striking about this portion of the *Meno* is what we might roughly describe as its empirical nature. One might say that what Socrates argues in this portion of the *Meno* is that the claim that virtue is a kind of knowledge, and so teachable, is incompatible with the facts on the ground.<sup>26</sup>

The empirical nature of the background beliefs that are found to be incompatible with the hypothesis is brought out in a number of ways throughout this passage. First, there is the puzzling passage at *Meno* 92b–c in which Anytus rejects the claim of the Sophists as teachers of virtue, explicitly disavowing any experience of them, but nevertheless claiming to know what sort the Sophists are. Consequently, Socrates' rejection of the Sophists' claim to being teachers of virtue is not affirmed until near the end of the “downward path” when Meno—whose experience with Gorgias is emphasized at the beginning of the dialogue—praises Gorgias for denying that he and other Sophists are teachers of virtue.<sup>27</sup> Second, the argument against Anytus' claim that the *kaloï kagathoi* are teachers of virtue rests on four specific counterexamples—Themistocles, Aristides, Pericles, and Thucydides, perhaps together with the probabilistic assumption that if anyone of the *kaloï kagathoi* could teach virtue it would be one of these four (see, for example, *Meno* 94c7–e2).<sup>28</sup> Third, the poets' claim to being teachers of virtue is rejected by a single counterexample—the self-contradictory claims of Theognis—without even a corresponding probabilistic assumption. Finally, there is the straightforward contingent nature of the argument structure: If it is possible to teach virtue, then virtue is actually being taught; but virtue is not actually being taught; so, it is not possible to teach virtue, and so virtue is not knowledge.<sup>29</sup>

Having confirmed the hypothesis that virtue is a kind of knowledge by an argument *to* the hypothesis at *Meno* 87c–89c, and then discrediting it at 89c–96d, Socrates is forced to attempt to resolve this tension. He does so by suggesting that the sub-argument in the argument *to* the hypothesis to the effect that nothing else is good other than knowledge is flawed. True belief, as well as knowledge, Socrates professes, is good. Consequently, virtue need not be a kind of knowledge as the argument *to* the hypothesis appeared to confirm. Rather, virtue may be true belief as well and so something that does not come to those who have it by teaching (or by nature) as the argument *from* the hypothesis confirmed, but by divine dispensation (*theia(i) moira(i)*). Of course, how we are to understand this resolution of the conflicting results of the method of hypothesis is a matter of considerable scholarly controversy—a controversy which we have no hope of resolving in this essay, but one which I maintain is illuminated by the argument of *Republic* 471c–502c.

### 13.3 The Structure of *Republic* 471c–502c

#### 13.3.1 *The First Stage: Identifying the Hypothesis (Republic 471c–473e)*

At *Republic* 471c–472b Glaucon suddenly<sup>30</sup> challenges Socrates to answer whether Kallipolis (as it has been described) is possible (*Rep.* 471e3–5). Before agreeing to take up this question Socrates first extracts a concession from Glaucon to the effect that he only be required to show how a city that closely approximates Kallipolis could come to be (*Rep.* 472e6–473b3).<sup>31</sup> Having secured Glaucon's agreement,

Socrates immediately proposes to reduce the question whether Kallipolis is possible to the question whether philosophy and political power coincide. He begins by indicating that there is one change—neither small nor easy—that would bring about the change to Kallipolis in existing cities (*Rep.* 473b4–c4) and follows this up by maintaining that unless this change takes place Kallipolis will never become possible (*Rep.* 473c11–e5). In the former passage Socrates testifies to the sufficiency of the change, while in the latter passage Socrates testifies to its necessity. The change that is neither small nor easy is variously described in the latter passage as “philosophers ruling as kings” (*Rep.* 473c11), “those now called kings and rulers philosophizing genuinely and adequately” (*Rep.* 473d1–2), and “political power and philosophy coinciding” (*Rep.* 473d2–3). For the remainder of this essay I will follow the lead of the last description and refer to the change necessary and sufficient for the possibility of Kallipolis as political power and philosophy coinciding.<sup>32</sup>

Notice that this introduction of the question whether Kallipolis is possible nicely parallels the introduction of the question whether virtue can be taught in the last third of the *Meno*. First, just as in the *Meno*, Socrates is portrayed as being forced to follow the lead of the interlocutor. Meno compels Socrates to pursue the question whether virtue can be taught against Socrates’ better judgment, just as Glaucon compels Socrates to pursue the question whether Kallipolis is possible. The language of compulsion is abundant in both passages. Second, as a result of being compelled to pursue a question against his better judgment Socrates extracts a concession from Glaucon, just as he had extracted a concession from Meno for being compelled to pursue the question whether virtue is teachable. The concession in the *Meno* is to be permitted to employ the method of hypothesis; here in the *Republic* the concession is more obscure,<sup>33</sup> but it is roughly that Socrates not be forced to show that Kallipolis should become possible in fact (*tô(i) ergô(i)*) in every detail in which it has been presented in theory (*tô(i) logô(i)*). Finally, and most significantly, Socrates reduces the question whether Kallipolis is possible to the question whether political power and philosophy coincide, just as in the *Meno* he had reduced the question whether virtue is teachable to the question whether virtue is knowledge. In the *Meno* the answer to the latter question is necessary and sufficient for the answer to the former; so here in the *Republic* the answer to the question whether political power and philosophy coincide is necessary and sufficient for the answer to the question whether Kallipolis is possible. If political power and philosophy coincide, Kallipolis is possible; if not, not. Moreover, just as in the *Meno* two theses compete for the title “hypothesis”—the thesis that virtue is knowledge and the biconditional that virtue is knowledge just in case virtue is teachable, so here in the *Republic* we have two theses that could plausibly be identified as hypotheses—the thesis that political power and philosophy coincide and the biconditional that political power and philosophy coincide just in case Kallipolis is possible. And again, just as in the *Meno* the former thesis receives the bulk of the attention in the remainder of the *Meno*, so here in the *Republic* the former thesis—that political power and philosophy coincide—receives the bulk of Plato’s attention in what follows.<sup>34</sup>

### 13.3.2 *The Second Stage, Part 1: The Argument to the Hypothesis (Republic 473e–487a)*

Let us then turn next to Plato's consideration of the thesis that political power and philosophy coincide. Glaucon immediately focuses on the hypothesis<sup>35</sup> as follows:

Socrates, after hurling a speech and statement like that at us, you must expect that a great many people (and not undistinguished ones either) will cast off their cloaks and, stripped for action, snatch any available weapon, and make a determined rush at you, ready to do terrible things. So, unless you can hold them off by argument and escape, you really will pay the penalty of general derision. (*Rep.* 473e6–474a4; Grube/Reeve transl.)

It is not immediately clear whether Glaucon is objecting to the biconditional or the thesis that political power and philosophy coincide, but Socrates' subsequent description of the argument to follow makes it clear that he takes Glaucon to be objecting to the latter. After securing Glaucon's promise to assist him in the argument that follows, Socrates lays out the structure of the argument.

I must try it, then, especially since you agree to be so great an ally. If we're to escape from the people you mention, I think we need to define for them who the philosophers are (*diorisasthai pros autous tous philosophous*) that we dare to say must rule. And once that's clear, we should be able to defend ourselves by showing that the people we mean are fitted by nature both to engage in philosophy and to rule in a city, while the rest are naturally fitted to leave philosophy alone and follow their leader.

This would be a good time to give that definition (*horizesthai*). Come, then, follow me, and we'll see whether or not there's some way to set it out adequately (*hikanōs*). Lead on. (*Rep.* 474b3–c7; Grube/Reeve transl.)

Here Plato makes clear that the argument that follows consists of two parts—first, an attempt to delineate the nature of genuine philosophy (*Rep.* 474c8–480a13), and second an argument that philosophy so understood includes the characteristics necessary and sufficient for genuine political power (*Rep.* 484a1–487a8). Thus, the argument that follows is aimed at confirming the truth of the hypothesis that political power and philosophy coincide. Moreover, the argument that follows conforms to the first of the two confirmation procedures we discussed above—"the argument to the hypothesis." Socrates is made to take the argument up to a hypothesis concerning the nature of philosophy that is adequate<sup>36</sup> (see *Phd.* 101e1) and then argue from that hypothesis back down to the original hypothesis that political power and philosophy coincide.

The portion of the argument concerned with coming to something adequate concerning the nature of philosophy has received considerable attention in the literature and I will not presume to add to it.<sup>37</sup> Rather, I will turn to the second portion of the argument conforming to the first procedure for confirming the hypothesis that political power and philosophy are one at *Republic* 484a1–487a8—a passage that has generated considerably less attention.

Socrates opens Book VI of the *Republic* by recapitulating the nature of philosophy "adequately" achieved in light of the arguments of the closing pages of Book V. Philosophy, Socrates maintains, is the ability to grasp what is always the same in all respects (*Rep.* 484b3–5). Philosophy, that is, is knowledge of Forms. Socrates

next asks whether philosophy so understood is the same as political power (*Rep.* 484b6–7), i.e., whether the higher hypothesis entails the original hypothesis.<sup>38</sup> He first points out that philosophy so understood is necessary for political power on the grounds that the knowledge that is philosophy is necessary to

establish here on earth conventions about what is fine or just or good, when they need to be established, or guard and preserve them, once they have been established. (*Rep.* 484d1–3; Grube/Reeve transl.)

And, next Socrates sets out to consider whether philosophy so understood is sufficient for the other necessary features of political power. As Socrates is made to put it:

Should we, then, make these blind people our guardians or rather those who know each thing that is and who are not inferior to the others, either in experience or in any other part of virtue? . . . Then shouldn't we explain how it is possible for someone to have both these sorts of qualities? (*Rep.* 484d5–485a2; Grube/Reeve transl.)

Plato here indicates that there are two necessary and jointly sufficient conditions for (genuine) political power: knowledge and virtue. He has already maintained that philosophy as delineated by means of the higher hypothesis is sufficient for the knowledge necessary for political power. So if he can show that philosophy is sufficient for virtue, he will have confirmed the original hypothesis (that political power and philosophy coincide) by deducing it from something adequate. And that is precisely what he goes on to do. After rehearsing yet again the nature of philosophy (*Rep.* 485a–c), Socrates goes on to argue that philosophy so understood entails (1) love of truth (*Rep.* 485c3–d5), (2) moderation (*Rep.* 485d6–e6), (3) liberality (*Rep.* 486a1–7), (4) courage (*Rep.* 486a8–b5), (5) justice, reliability, gentleness (*Rep.* 486b6–13), (6) fast-learning (*Rep.* 486c1–6), (7) memory (*Rep.* 486c7–d3), and (8) measure and calm (*Rep.* 486d4–12). Socrates concludes:

Well, then, don't you think the properties we've enumerated are compatible with one another and that each is necessary to a soul that is to have an adequate and complete grasp of that which is?

They're all completely necessary.

Is there any objection you can find, then, to a way of life that no one can adequately follow unless he's by nature good at remembering, quick to learn, high-minded, graceful, and a friend and relative of truth, justice, courage, and moderation?

Not even Momus could find one.

When such people have reached maturity in age and education, wouldn't you entrust the city to them and to them alone? (*Rep.* 486e1–487a8; Grube/Reeve transl.)<sup>39</sup>

At this point, then, Socrates has reduced the question whether Kallipolis is possible to the question whether philosophy and political power coincide and confirmed a positive answer to the latter question by deducing it from a higher adequate hypothesis concerning the nature of philosophy. The structure of the argument during this stretch of the *Republic* then nicely parallels the argument structure of the *Meno* from 87b2–89c4 which reduced the question whether virtue is teachable to the question whether virtue is knowledge and then went on to deduce a positive answer to the latter question from the higher adequate hypothesis that virtue is good. But before we turn to the next portion of the argument in the *Republic*, we should notice that it

is not merely the structure of the present argument that parallels the argument in the *Meno*, but the substance of the argument as well.

Remember that the argument in the *Meno* is aimed at establishing that knowledge of some sort (knowledge of the good or perhaps the knowledge possessed by the philosopher) is necessary and sufficient for virtue. Clearly, this is the topic on the table here in the *Republic*. The argument from 485a–486d amounts to an argument for the thesis that knowledge of a certain sort (knowledge of the Forms or knowledge of the philosopher) is sufficient for virtue. But the virtue or virtues necessary for knowledge in this passage appear unlike the virtues discussed at the end *Republic* IV. They appear more propaedeutic, incomplete or imperfect. The discussion at the end of *Republic* IV suggests that complete justice can only be attained by an individual who has the knowledge of the wisdom-loving part (of the city and the soul), for example. The knowledge discussed here in *Republic* VI, the knowledge of the philosopher, then, turns out to be both necessary and sufficient for genuine virtue.<sup>40</sup> But this is precisely the thesis defended at *Meno* 87b2–89c4. Rather than pursuing this further it is enough for our current purposes to note the similarity of subject matter between these two arguments. It is difficult to imagine that Plato does not have the last third of the *Meno* in mind as he composes this portion of the *Republic*. And so we would do well to have it in mind as well when we turn to the next portion of the argument in the *Republic*.

### 13.3.3 *The Second Stage, Part 2: The Argument from the Hypothesis (Republic 487a–502c)*

Having confirmed the hypothesis that philosophy and political power coincide via an argument *to* the hypothesis from a higher hypothesis, Socrates is confronted by Adeimantus as follows:

No one would be able to contradict the things you've said, Socrates, but on each occasion that you say them, your hearers are affected in some such way as this. They think that, because they're inexperienced in asking and answering questions, they're led astray a little bit by the argument at every question and that, when these little bits are added together at the end of the discussion, great is their fall, as the opposite of what they said at the outset comes to light. . . . Yet the truth isn't affected by this outcome. I say this with a view to the present case, for someone might well say now that (*logô(i) men*) he's unable to oppose you as you ask each of your questions, yet he sees (*ergô(i) de*) that of all those who take up philosophy—not those who merely dabble in it while still young in order to complete their upbringing and then drop it, but those who continue in it for a longer time—the greatest number become cranks, not to say completely vicious, while those who seem completely decent are rendered useless to the city because of the way of life you recommend. (*Rep.* 487b1–d5; Grube/Reeve transl.)

Notice that Adeimantus does not here object to the preceding argument. Indeed, he grants that he is unable to oppose it. But he denies that the conclusion is true. The conclusion that philosophy and political power coincide is simply contradicted by the facts on the ground. Counterexamples, he suggests, are almost too numerous to mention. Most philosophers are vicious and so not genuine political rulers. The rest

are useless. Adeimantus is not here reiterating Glaucon's objection at 475d–e. He is not misunderstanding what philosophy is and so misidentifying who the philosophers are. Rather, he maintains that the philosophers, as Socrates understands them, are in fact either vicious or useless, not genuine rulers. The empirical nature of Adeimantus' argument is underscored by the *logos/ergon* distinction at 487c5–6 which is reiterated throughout the remainder of the passage.<sup>41</sup> (It is also, by the way, in violation of Glaucon's concession at the beginning of the argument [*Rep.* 472e6–473b3]). Adeimantus does not dispute the *logos* on behalf of the conclusion that philosophy and political power coincide, but he does take the conclusion to be inconsistent with the *ergon*. The empirical nature of Adeimantus' challenge recalls the empirical nature of the argument that discredited the hypothesis that virtue is knowledge at *Meno* 89c–96d. There Socrates maintained that the hypothesis that virtue is knowledge was refuted by the fact that there were no teachers of virtue. Here, Adeimantus maintains that the hypothesis that philosophy and political power coincide is refuted by the fact that philosophers are either vicious or useless, not genuine rulers. Here in the *Republic*, then, begins that portion of the argument in the *Meno* that we called the argument *from* the hypothesis.

The parallel with the *Meno* continues. In the *Meno* Socrates accepts the facts on the ground that evidently contravene the hypothesis that virtue is knowledge. Indeed, he is the one who puts them forward. Here in the *Republic* Socrates also accepts the facts on the ground that evidently contravene the hypothesis that philosophy and political power coincide. He concedes that philosophers, as he has defined them, are useless and vicious. He immediately responds to Adeimantus' challenge by conceding that what Adeimantus has said is true (“they seem to me to speak the truth”; *Rep.* 487d10 Grube/Reeve transl.), and reiterates this concession at least two more times (*Rep.* 489b3, d5).<sup>42</sup> Nevertheless, he does not concede that the truth of the claim that philosophers are vicious or useless contravenes the hypothesis that philosophy and political power coincide. He does not concede that the facts on the ground are inconsistent with the consequences of the hypothesis that philosophy and political power coincide.

The parallel with the *Meno* finally collapses. In the *Meno* Socrates accepts the fact that there are no teachers of virtue and agrees that this fact contravenes the hypothesis that virtue is knowledge. Here in the *Republic*, Socrates accepts the fact that philosophers are vicious or useless but denies that this fact contravenes the hypothesis that philosophy and political power coincide. How can this be? How can Socrates accept the fact that philosophers are either vicious or useless and the hypothesis that philosophy and political power coincide?

Space does not permit a detailed analysis of the argument that follows. Fortunately a detailed analysis of the argument is unnecessary given our focus on the philosophical method being employed in this portion of the *Republic* as opposed to its philosophical content. The argument falls into three distinct parts. First, an account of why the decent philosophers are useless (*Rep.* 487d–489d); second, an account of why most philosophers are vicious (*Rep.* 489d–496a); and third, an account of how it nevertheless remains possible even given these facts for philosophy and political power to coincide (*Rep.* 497a–502c).



Let us first look briefly at the account of why the decent philosophers are useless. Socrates is made to appeal to the image of the ship. We are to think of Athens (or any other Greek *polis* in which philosophers are either useless or vicious) as a ship with the ship-owner standing for the Demos.<sup>43</sup> The ship-owner who is described as bigger and stronger than everyone on board, but hard of hearing, short-sighted, and ignorant of sea-faring, is continuously implored by individual sailors to be permitted to steer and rule the ship. The sailors are described as quarreling with each other, each thinking he or she should rule, never having learned the *technê* of steering, nor being able to point to anyone who taught them the *technê* nor to a time at which they learned it. Indeed, they claim it is not teachable and are ready to kill anyone who maintains that it is. Moreover, they call those who are clever at persuading the ship-owner to let them rule navigators, captains, and those who know ships, dismiss anyone else as useless, and do not understand what a true captain should care about, i.e., the seasons, the sky, the stars, the winds and everything appropriate for his *technê*. Finally, they call the true captain a star-gazer, a babbler, and useless.

It is difficult as we read the description of the sailors on this ship not to think of Anytus, who was clever at persuading the Demos that he should rule, despite lacking the *technê* to rule, not being able to point to anyone who taught the *technê* (other, I suppose, than any Athenian gentleman) nor anytime at which he learned it, and who was not only ready to, but did, kill someone who maintained that the *technê* was teachable—Socrates.

Having described the ship in terms reminiscent of Anytus, Socrates is made to recount how the image of the ship explains the uselessness of philosophers. It is true he repeats (*Rep.* 489b3) that the philosophers are useless in present cities, just as the genuine captain is useless on the ship. But philosophy is not to blame (*aitiasthai*) for the philosophers' uselessness. Rather, it is "those who don't make use of the philosophers" who are to blame—the Demos and/or the demagogues. The Demos which is already short-sighted is blinded by the demagogues to the utility of philosophy. It is not the nature (*phusin*) of a genuine captain to beg the sailors to rule, nor of the genuine ruler, or philosopher, to beg the Demos to rule.<sup>44</sup> Consequently, it is not easy for the genuine philosopher to rule in the present climate. Indeed, it would be surprising if he or she did. Socrates is made to conclude this portion of the argument as follows:

And haven't we explained why (*tên aitian*) the decent ones are useless? (*Rep.* 489d7–8; Grube/Reeve transl.)

Before moving on to the next portion of the argument we should note how the current argument is supposed to go. Socrates does not deny the truth of the claim that in the current climate philosophers are useless. He does not object that Adeimantus has misunderstood what he means by philosophy or philosophers. Rather he explains by means of the image of the ship that philosophy is not the cause (*tên aitian*) of this truth. It is not a consequence of the nature of philosophy that philosophers in current cities are useless. Rather it is a consequence of the blindness of the Demos and the obfuscating practices of the demagogues. As a result, evidently, Plato thinks the

hypothesis escapes refutation. The hypothesis that philosophy and political power coincide is compatible with philosophers being useless in current cities because the nature of philosophy is not the cause (*hê aitia*) of philosophers' uselessness.

A similar point appears to be the theme of the next portion of the argument—as Socrates is made to put it:

Then, do you next want us to discuss why it's inevitable that the greater number are vicious and to try to show, if we can, that philosophy isn't responsible (*aitia*) for this either? (*Rep.* 489d1–e1; Grube/Reeve transl.)

Again, Socrates will concede the facts on the ground (viz. that the majority of philosophers are vicious) although this time not literally in the way that Adeimantus presents them,<sup>45</sup> but denies that they are inconsistent with the consequences of the hypothesis that philosophy and political power coincide because philosophy is not the cause of philosophers' viciousness. The argument begins with a recapitulation of the nature of philosophy or the philosophic nature (*Rep.* 489e3–490e1),<sup>46</sup> followed by a description of the argument to follow:

We must now look at the ways in which this nature is corrupted, how it's destroyed in many people, while a small number (the ones that are called useless rather than bad) escape. After that, we must look in turn at the natures of the souls that imitate the philosophic nature and establish themselves in its way of life, so as to see what the people are like who thereby arrive at a way of life they are unworthy of and that is beyond them and who, because they often strike false notes, bring upon philosophy the reputation that you said it has with everyone everywhere. In what ways are they corrupted? (*Rep.* 490e2–491a6; Grube/Reeve transl.)

Socrates is here made to distinguish two parts of the subsequent argument—a part devoted to how genuine philosophy or the philosophic nature is corrupted (*Rep.* 491a7–495b7) and a part devoted to explaining how those who do not have a genuine philosophic nature appear to be philosophers and provide a false reputation to genuine philosophy (*Rep.* 495b8–496a10). The longer part can be summarized as follows.

The passage begins with Socrates maintaining that the qualities or natural abilities that he had argued were necessary for philosophy back at 485c–d are rare and when combined with a good education lead to complete virtue but when combined with a bad education will lead to vice “unless some god happens to come to its rescue” (*Rep.* 491a7–492a5).<sup>47</sup> Socrates next is made to explain how traditional education corrupts those with the necessary qualities for philosophy (*Rep.* 492a5–493a5) and then how sophistic education corrupts those with this nature as well (*Rep.* 493a6–494a10). This is followed with an account of the corruptive influence of family, friends and other flatters (*Rep.* 494a11–495a3).<sup>48</sup> The argument concludes:

Do you see, then, that we weren't wrong to say that, when someone with a philosophic nature is badly brought up, the very components of his nature—together with the other so-called goods, such as wealth and other similar advantages—are themselves in a way the cause (*aitia*) of his falling away from the philosophic way of life? (*Rep.* 495a4–8; Grube/Reeve transl.)

Like the argument *from* the hypothesis in the *Meno*, then, Socrates goes through the traditional modes of education—one’s elders and the Sophists<sup>49</sup>—and finds them wanting. In fact, here in the *Republic* they are not simply unsuccessful, they are positively harmful.<sup>50</sup> But unlike the argument in the *Meno*, Socrates does not concede that these facts contravene the hypothesis that philosophy and political power coincide. Rather he argues that the qualities and natural abilities necessary for philosophy when combined with traditional forms of education are the cause of the viciousness of philosophers. The nature of philosophy is not the cause.

Next, Socrates is made to explain that in circumstances like these philosophy is left deserted and those lacking the qualities and natural abilities necessary for philosophy move in and take up philosophy. These are the ones, Socrates says,

who are responsible for the reproaches that you say are cast upon philosophy by those who revile her, namely, that some of those who consort with her are useless, while the majority deserve to suffer many bad things (*Rep.* 495c4–6; Grube/Reeve transl.)

for their thoughts and beliefs

are properly called sophisms, things that have nothing genuine about them or worthy of being called true wisdom (*Rep.* 496a7–9; Grube/Reeve transl.).

Again, Socrates concedes that the majority of those who practice philosophy in existing cities are vicious, but this time he denies that those who practice philosophy in existing cities are genuine philosophers—for they lack the qualities and natural abilities necessary for genuine philosophy. Again, the cause of the viciousness of philosophers is not philosophy, but traditional education—on the one hand, corrupting those who possess the qualities and natural abilities necessary for the genuine philosophical life and leaving room for those who fail to possess these qualities to take up the genuine philosophical life, on the other. Socrates is made to conclude this portion of the argument by explaining that the few decent and useless ones who escape the corrupting influence of traditional education do so roughly by escaping notice (*Rep.* 496a11–e2).

Having established that the corrupting influence of traditional education causes the viciousness of most philosophers, that the blindness of the Demos and/or the demagogues causes the uselessness of the rest, and that the nature of philosophy causes neither in the current climate, it remains for Socrates to show that in this climate it is possible for philosophy and political power to coincide. First (*Rep.* 497a–498c4), Socrates is made to explain the changes in traditional education that would mitigate its corrupting influence. The key change is reserving genuine philosophical study to old age. Socrates is made to elaborate the nature of this education—the kind that would lead to philosophy and political power coinciding—throughout the rest of Books VI and VII (esp. 518b–540c). In the meantime, we must depend on chance or divine inspiration (*ek tinos theias epipnoias*) for the coincidence of philosophy and political power (*Rep.* 499a10–e6). While such a coincidence is rare indeed, Socrates maintains “it cannot be reasonably maintained . . . that either of these things is impossible” (*Rep.* 499c2–3; Grube/Reeve transl.). The remainder of the passage is devoted to explaining that the rarity of this coincidence

as well as the consequent failure of the majority to understand the genuine nature of philosophy accounts for the difficulty of persuading the majority of the value of the coincidence of philosophy and political power. And so, Socrates draws the argument aimed at addressing the third wave begun back at 471c to a close as follows: “Then we can now conclude that this legislation is best, if only it is possible, and that, while it is hard for it to come about, it is not impossible” (*Rep.* 502c5–7; Grube/Reeve transl.).<sup>51</sup>

### 13.4 The Salient Difference and Two Kinds of Consequences

One of the many puzzles surrounding the method of hypothesis is making sense of Socrates’ description in the *Phaedo* (101d3–5) of the procedure I have been calling the argument *from* the hypothesis. Socrates maintains that one should examine “the consequences that follow from [the hypothesis to see whether they] agree with one another or contradict one another (*ta ap’ ekeinês hormêthenta skepsaio ei soi allêlois sumphônei ê daiphônei*.)” What is puzzling about this is that on a natural understanding of “consequences” and “agreement” this procedure amounts to examining whether the hypothesis is self-contradictory. Only a self-contradictory proposition can have deductive consequences that are inconsistent. But while one certainly would want to discredit self-contradictory hypotheses, this hardly looks like a very productive procedure. The odds of proposing a hypothesis that is self-contradictory are pretty slim.<sup>52</sup> Now there are well-known issues concerning the correct understanding of agreement in this context,<sup>53</sup> but our examination of the application of the procedure of the argument *from* the hypothesis at *Republic* 487a–502c suggests we may begin to make some headway by focusing on a special way of reading “consequences.”

Readers familiar with the *Republic* will recall an earlier passage in which the nature of consequences plays an important role. At the beginning of Book II, Glaucon presses Socrates to defend the claim that justice is a good welcomed for its own sake as well as its consequences. Again on a natural understanding of “consequences” it is common to read the distinction between welcomed for its own sake and welcomed for its consequences as a distinction between intrinsic and extrinsic value. Nevertheless a variety of scholars have raised difficulties for understanding the distinction in this way and consequently have proposed understanding the distinction as a distinction between two kinds of consequences.<sup>54</sup> The distinction is controversial, but it is roughly a distinction between the natural or necessary consequences of justice, for example, and the artificial, contingent, or societal consequences of justice. The idea is that Socrates is challenged to show not only that justice is beneficial in the current climate, given the difficulty of avoiding detection for injustice and the need to appear just to reap various societal rewards like wealth, political power, and a pleasant afterlife, but that justice would be beneficial even in different circumstances, in which detection for injustice was easier to avoid (Gyges’ ring; *Rep.* 359c–360b) or one’s injustice was rewarded with wealth, political power, and good marriage (choice of lives; *Rep.* 360e–362c). Socrates must

show that benefits flow directly from the nature of justice itself, and not from justice together with other artificial or contingent matters of fact. We might put the distinction as follows: Socrates must show that justice itself is the cause (*aitia*) of the benefits of justice.<sup>55</sup> He must show that it is not the case that all of the benefits of justice are caused by justice together with contingent features of the current climate.

Keeping this distinction in mind, let us return to the difference between the argument *from* the hypothesis in the last third of the *Meno* and in *Republic* 471c–502c. In the *Meno*, Socrates is made to propose that the hypothesis that virtue is a kind of knowledge had the consequence that virtue has teachers (and learners), given the background beliefs that all knowledge is teachable and everything teachable has teachers (and learners). This consequence of the hypothesis, however, was found to be inconsistent with the facts on the ground. In the current climate, no teachers (or learners) of virtue are to be found. Consequently, the hypothesis that virtue is knowledge is discredited because—to use the language of the *Phaedo*—the consequences of the hypothesis are not in harmony (*diaphônei*). Analogously, in the *Republic*, Socrates is evidently proposing that a consequence of the hypothesis that philosophy and political power coincide is that philosophers are useful and virtuous, perhaps given the background beliefs that genuine political power is useful and an argument through the nature of philosophy that genuine political power is necessary and sufficient for genuine virtue. This consequence of the hypothesis, however, appeared to be inconsistent with the facts on the ground. In the current climate philosophers are neither useful nor virtuous. Socrates is made to preserve his hypothesis in the *Republic*, however, by arguing that philosophy is not the cause (*aitia*) of the facts on the ground. The uselessness and viciousness of philosophers in the current climate is not a consequence of the nature of philosophy, but a consequence of the nature of philosophy together with various other artificial and contingent matters. There is a sense, then, in which the “consequences” of the hypothesis that philosophy and political power coincide are inconsistent, but Socrates argues that one of those “consequences” is not a natural or necessary consequence of the nature of philosophy.

If something like this is on the right track,<sup>56</sup> we should notice that the natural but uncharitable reading of the *Phaedo* passage turns out to be correct but perhaps not so uncharitable. The argument *from* the hypothesis consists in examining whether the natural consequences of the hypothesis are consistent or inconsistent. It remains true that only a self-contradictory hypothesis would be effectively ruled out by this procedure, but it may now appear more plausible to suppose that it will be productive to show that a hypothesis is not self-contradictory. What one must do is examine whether the consequences of the hypothesis are inconsistent, i.e., examine whether the effects of the Form alone are inconsistent (“without the images used in the previous section, using Forms themselves and making its investigation through them”; *Rep.* 510b7–9), and explain how apparent natural consequences of the hypothesis are in fact artificial or contingent (see *Rep.* 534b8–c3). It is only when one can successfully complete the argument *from* the hypothesis in this way (after having completed the argument *to* the hypothesis up to the unhypothetical first principle of everything), that one can be genuinely said to have knowledge.

### 13.5 Conclusion

In closing let us return to the argument of the last third of the *Meno* with which this essay began. The similarities between this argument and *Republic* 471c–502c should be apparent. The allusions in the *Republic* to the *Meno* are abundant. Both passages begin with Socrates being compelled to pursue a question apparently against his will. In both passages Socrates is made to accede to this compulsion after extracting a concession from his interlocutor. Both passages make meaningful use of “divine dispensation.” Anytus appears to be alluded to in the description of the sailors who manage to gain control of the ship of state despite their lack of expertise. *Meno* even may be alluded to in the description of those with the natural abilities necessary for philosophy who get corrupted by traditional education. And of course both passages are concerned with identical subject matter—the relationship among knowledge (or wisdom or philosophy), virtue (or political power), and education.

These similarities are reinforced by the parallel argument structure—a structure which conforms to the method of hypothesis as elaborated in the *Meno*, the *Phaedo*, and *Republic*. Both passages begin by reducing the question Socrates is compelled to pursue to another question from which the answer to the original question can be deduced. In the *Meno*, Socrates is made to reduce the question whether virtue is teachable to the question whether virtue is a kind of knowledge, and in the *Republic* Socrates is made to reduce the question whether Kallipolis is possible to the question whether philosophy and political power coincide. These portions of the arguments correspond to the first stage of the method of hypothesis—the reduction stage—in which one (a) seeks to identify a hypothesis from which an answer to the question can be derived and (b) shows how the truth of the hypothesis entails the answer to the question. Indeed, both dialogues focus on the first of these procedures, taking the second procedure to be relatively straightforward.

Again, in both passages, Socrates is made to turn next to the second or confirmation stage of the method of hypothesis—taking up the argument *to* the hypothesis first. In the *Meno* Socrates confirms the hypothesis that virtue is knowledge by deriving it from the higher hypothesis that virtue is good and in the *Republic* Socrates confirms the hypothesis that philosophy and political power coincide by deriving it from the higher hypothesis that philosophy is the knowledge of the Forms. Having confirmed the respective hypotheses by means of a primarily theoretical argument (*tô(i) logô(i)*), both dialogues turn to the confirmation process concerned with the facts on the ground (*tô(i) ergô(i)*)—the argument *from* the hypothesis. In the *Meno* the hypothesis that virtue is a kind of knowledge is discredited on the grounds that its consequence that there are teachers (and learners) of virtue is inconsistent with the fact that there in fact are no teachers of virtue. In the *Republic*, the hypothesis that philosophy and political power coincide appeared to be discredited on the grounds that its consequence was inconsistent with the facts that philosophers are in fact useless or vicious. But Socrates is made to defend his hypothesis in the *Republic* against this attack on the grounds that philosophy is not the cause (*aitia*) of the uselessness and/or viciousness of philosophers in the current climate.

The difference between the two passages is significant, but it should not obscure the similarity of method. In both dialogues Plato exhibits an application of the method of hypothesis. But the difference should affect our understanding of the application in the *Meno*, for the difference highlights a feature lacking in the *Meno*. While both applications appeal to facts on the ground that appear to discredit the hypothesis as part of the confirmation process, the *Republic* makes clear that appeal to such facts is insufficient for abandoning the hypothesis. In addition to providing the conflicting facts on the ground one must determine whether those facts are a natural consequence of or caused by the hypothesis itself (or the Forms or concepts used to deduce the hypothesis). When they are, the hypothesis must be abandoned; when they are not, the hypothesis can be retained. And when the hypothesis can be retained against all such conflicting facts it has been confirmed (at least with respect to the argument *from* the hypothesis). Of course, no such determination has been made in the *Meno*. Socrates has been made to present only the conflicting facts on the ground. He has not determined nor even attempted to determine the cause of those facts. Consequently, the results of the application of the method of hypothesis in the *Meno* are untested and incomplete (at least by the lights of the *Republic*) and Socrates is correct to profess his ignorance of the nature of virtue at the end of the *Meno* (100b4–6). We too would be wise to hesitate in taking the argument in the last third of the *Meno* as endorsing a particular understanding of the nature of virtue. But at *Republic* 471c–502c progress appears to have been made along the longer road to the unhypothetical first principle of everything.<sup>57, 58</sup>

## Notes

1. The *Meno* can be seen as falling into three parts: (1) an attempt to answer the “What is virtue?” question (70a–79e), (2) a methodological digression, containing Meno’s paradox, the theory of recollection, the conversation with the slave, and an argument for the immortality of the soul (80a–86c), and (3) the discussion concerning the teachability of virtue (86c–100a).
2. Although arguments in which Socrates is made to argue for both *p* and not-*p* are not as common as one might think, perhaps the best examples of such arguments can be found in the *Cratylus*, where Socrates appears to argue first against conventionalism and then against naturalism, and especially in the *Parmenides*.
3. By the “Socratic dialogues” I mean those dialogues which have also been described as “early” or “aporetic”: *Apology*, *Charmides*, *Crito*, *Euthydemus*, *Euthyphro*, *Gorgias*, *Hippias Major*, *Hippias Minor*, *Ion*, *Laches*, *Lysis*, and *Protagoras*.
4. This has often been taken as the first step in the rejection of Socratic intellectualism in Plato’s so-called middle dialogues. For some reason to doubt this change or development on Plato’s part, see Carone (2001).
5. Both captured by the same Greek word—*didakton*. See Scott (2006, p. 162) for a plausible way to avoid the equivocation at least in its crudest form.
6. See, for example, Cornford (1957, p. 245), and Bedu-Addo (1984, pp. 10–14).
7. See Benson (2003 and 2006).
8. For an argument that Plato endorses the method of hypothesis see Benson (2003).
9. After composing some early drafts of this essay I reread Nettleship’s lecture IX entitled “Philosophy and the State: [*Rep.* 471c–502c]” (Nettleship 1925, pp. 184–211). This is one of the few examples in the literature of a treatment of the entire stretch of argument from



471c–502c. Most scholars focus on the arguments at the end of *Republic* V and then refocus at 502c and especially on the analogies of the Sun, Line and Cave. I was happy to discover that in the main my reading of the structure of the argument over this stretch of text was in sympathy with Nettleship's. He does not, however, see (or at least maintain) the parallel with the last third of the *Meno* nor the application of the method of hypothesis.

10. I defend the view that Plato recommends roughly the same method of philosophical investigation in these three dialogues in Benson (2006). The present essay is an additional piece of that argument. Few commentators would fail to take seriously Plato's argument from 471c–502c, and yet if the present essay is correct, Plato employs his method of hypothesis in pursuing this argument. While I am inclined to think that Plato's views concerning the method of hypothesis and dialectic undergo a development in the *Meno*, *Phaedo*, and *Republic*, nothing I say in this essay depends on that inclination. So, on this issue as well, I leave to one side issues concerning developmentalism.
11. For the phrases "an argument *to* the hypothesis" and "an argument *from* the hypothesis" see Benson (2003). These two procedures correspond roughly to what has in the literature sometimes been called "the upward path" and "the downward path." See esp. Robinson (1953, pp. 162–77). One of the goals of my recent work (including the present essay) is to put a more precise formulation on the nature of these two procedures.
12. For the longer and shorter road, see *Republic* 504a4–d3 and for the unhypothetical first principle, see *Republic* 510b5–6, 511b6–7, and 533c8–d1.
13. See Benson (2003, 2005, 2006). Differences between and among these essays and the present one indicate my own development in understanding Plato's dialectical method. The present essay supersedes the previous ones, but I do not anticipate that it will be my final judgment on the matter. As is the case with every aspect of Plato's view there are always further passages to consider, literature to read, and puzzles to be teased out.
14. For a defense of the claim that Socrates' reluctance here does not amount to Plato's reluctance to endorse the method of hypothesis see Benson (2003, pp. 6–12).
15. Socrates' precise words are "...allow the question—whether virtue comes by teaching or some other way—to be examined by means of hypothesis (*ex hupotheseōs auto skopeisthai*). I mean by hypothesis what the geometers often do in dealing with a question put to them (*hōsper hoi geometrai pollakis skopountai, epeidan tis erētai autous*); for example. . ." (*Meno* 86e2–5; Lamb transl.). In maintaining that this is the first time in the dialogues that the method of hypothesis is explicitly referred to, I leave open the possibility of which I am rather dubious that it is practiced earlier (as Kahn [1996, pp. 184, 196] and perhaps Weiss [2001, p. 114 n. 78] appear to think). Moreover, in assuming that the so-called Socratic dialogues were composed earlier than the *Meno*, I continue to leave open the question of Platonic development. Whether Plato's philosophical position ever developed or not, he did not write all of the dialogues on the same day. The totality of the evidence suggests that the so-called Socratic dialogues were written relatively early and probably earlier than the *Meno*—but in the end nothing in this essay hangs on this. The literature on the method of hypothesis is enormous. For a bibliography see Benson (2003).
16. The entailment need not be so immediate as it is in the *Meno*. The entailment from the truth of the hypothesis that Forms exist (or the theory of Forms) to the conclusion that the soul is immortal at *Phaedo* 100b–107b is hardly immediate. See also *Phaedo* 76d7–77a5 where Plato suggests Forms exist just in case souls exist before birth.
17. See note 24 below.
18. Although, as I will maintain in a moment, he does attempt to illustrate by example the confirmation process—however unsuccessfully.
19. Ironically, he *illustrates* the reduction stage in the *Phaedo*.
20. By "choosing the hypothesis" I mean choosing, for example, that virtue is a kind of knowledge, or that virtue is not a kind of knowledge.
21. Dialectic is sometimes viewed by Plato as the culmination of a method aimed at complete knowledge or understanding and sometimes as merely the method aimed at complete

- knowledge or understanding. Understood in the former way, any attempt at confirming a hypothesis that falls short of either confirmation procedure would not amount to an application of dialectic. In this case, Plato might prefer to think of such applications as applications of the method of hypothesis (more broadly construed) or of dianoetic (construed as any application of the method of hypothesis that falls short of dialectic). Nevertheless, such applications would be instances of dialectic understood in the latter way, i.e., as the method aimed at complete knowledge or understanding whether successful or not. In this case, dianoetic should be restricted to any method that employs only the first stage of the dialectic method.
22. As is well known the literature on the Divided Line is vast and diverse; the literature on Plato's educational curriculum in *Republic* VII is only slightly less so.
  23. See Section 13.4 for a suggestion concerning the way the testing of the hypothesis is a priori. For a somewhat longer defense of this rough outline of the dialectical method see Benson (2006).
  24. Considerable amount of ink has been spilled concerning whether *the* hypothesis here is “virtue is teachable just in case virtue is knowledge” or “virtue is knowledge.” As I suggest in Benson (2003) this issue need not detain us since it is clear in the *Meno* that Socrates is willing to designate a variety of propositions as hypotheses. (For some of the literature involved in this debate see Benson [2003, nn. 53, 54].) We might speculate that the appropriateness of designating one or the other of these sorts of propositions as the hypothesis varies with their respective clarity and credibility. For example, when the connection between the two sides of the biconditional is relatively straightforward as it is in the *Meno* (and as we will see in the *Republic*), the proposition representing one side of the biconditional is more appropriately seen as the hypothesis since it is the proposition that will need to undergo the more elaborate explanation and defense. On the other hand, when the biconditional itself is hardly straightforward and so in need of explanation and defense as is the biconditional in the *Phaedo*—roughly that Forms exist just in case the soul is immortal, it is the biconditional that is appropriately designated as the hypothesis. In fact, however, in the *Phaedo* it is the proposition that Forms exist that is designated as the hypothesis. Scott (2006, pp. 221–224) has plausibly rejected the biconditional interpretation. He admits, however, that “at the level of textual detail, both interpretations have their difficulties” (p. 224). Scott plausibly suggests that one of the key features of the hypothesis is its provisionality and only the proposition that virtue is knowledge is provisional in the *Meno*. This fits well with the suggestion above that which of the two propositions is dubbed the hypothesis varies with its relative clarity or credibility.
  25. As Scott (2006, p. 157) sees, the argument *to* the hypothesis here in the *Meno* is insufficient for another reason as well. The argument *to* the hypothesis fails to go all the way up to the Form of the Good or the unhypothetical first principle of everything as is required by the method as elaborated in the *Republic*.
  26. For a longer defense that this portion of the *Meno* corresponds to the argument *from* the hypothesis see Benson (2003, pp. 21–25).
  27. Even here it is hardly affirmed confidently, since Meno goes on to claim he cannot say whether the Sophists are teachers of virtue, sometimes he thinks they are, and sometimes he thinks they are not (*Meno* 95c5–8). I might note in passing that being a teacher of virtue is not something Gorgias denies in the *Gorgias*.
  28. It may be worth noting that throughout this discussion of these counterexamples Anytus repeatedly makes use of perception verbs: he had heard (*akêkoas*) that Themistocles taught his son Cleophantus to be a good horseman, but he had never “heard (*akêkoas*) anyone, young or old, say that Cleophantus, the son of Themistocles, was a good and wise man at the same pursuits as his father” (*Meno* 93d9–e5; Grube transl.); he had seen (*hora(i)s*) the kind of man Aristides' son was; and he had heard (*akoê(i)*) that Thucydides had provided the best wrestling teachers for his sons. See also Scott (2006, pp. 187–192) who plausibly defends the view that Plato is serious about the virtue (at least *qua* true belief) of these political leaders.
  29. On the empirical focus of this argument see Scott (2006, pp. 177–78).
  30. Despite the fact that the question has been in the background since *Rep.* 458a–b (and even since 450c–d).

31. Socrates here draws for the first time in this passage the distinction between *logos* and *ergon* which will get reiterated throughout the passage.
32. See Nettleship (1925, p. 186) who writes: “the union of political power and philosophical insight.”
33. See Halliwell (1993, p. 198) who writes that Plato’s “contention here is a source of difficulty.”
34. Why Plato fails to take the biconditional in the *Republic* to require explanation and defense is less clear than in the *Meno*. The biconditional in the *Meno* is a substitution instance of a relatively endoxic principle that something is teachable just in case it is knowledge; see, for example, *Protagoras* 361a3–c2 and *Timaeus* 51e2. Perhaps, our inability to recognize the obviousness of the biconditional in the *Republic* is a consequence of the normative and modal aspects of Plato’s discussion here. The biconditional might be more accurately characterized as follows: a genuine city is possible just in case it is possible for genuine political power and genuine philosophy to coincide (where the notion of genuineness is meant to capture something like a normative ideal). Characterized in this way, what may seem most problematic is the claim that genuine political power and genuine philosophy could coincide, not that a genuine city is possible just in case this claim is true. For simplicity I will continue to refer to the *Republic*’s biconditional as “Kallipolis is possible just in case political power and philosophy coincide.”
35. See the dispute over *exoitō* at *Phaedo* 101d3: Robinson (1953), Gallop (1975, p. 235), Kanayama (2000, pp. 76–8), Kahn (1996, p. 318 n. 35), and Dancy (2004, pp. 298–9).
36. See also *Republic* 485a6.
37. The literature devoted to the end of *Republic* V is of course enormous. As a suggestion, the entire passage may conform to a case of the first confirmation stage on the higher hypothesis concerning the nature of philosophy. After coming to an adequate account of the nature of philosophy at 475b8–c8 to the effect that philosophy is the desire for all wisdom and learning as a result of somewhat suspect *epagōgē*, the account is “attacked” or “clung to” by Glaucon on the grounds that “many strange people will be philosophers.” Socrates responds to this by seeking a higher hypothesis concerning the nature of wisdom or knowledge by which the hypothesis that philosophy is the desire for all such wisdom and knowledge can be confirmed.
38. Throughout these pages Socrates is made to put the argument in terms of the nature of a philosopher, rather than philosophy, but we have seen since 473c–e that the question at issue during these pages in the *Republic* is variously (and presumably equivalently) put as whether philosophers are rulers or whether philosophy and political power are one.
39. We can put the structure of this argument as follows:
  - [1] Philosophy  $\Leftrightarrow$  sophia (V.474b–480a)
  - [2] Political power (PP)  $\Leftrightarrow$  sophia and virtue (VI.484d5–485a2)
  - [3] PP  $\Rightarrow$  sophia (from [2])
  - [4] Sophia  $\Rightarrow$  philosophy (from [1])
  - [5] PP  $\Rightarrow$  philosophy (from [3] & [4])
  - [6] Philosophy  $\Rightarrow$  sophia (from [1])
  - [7] Sophia  $\Rightarrow$  virtue (VI.485a–486d)
  - [8] Philosophy  $\Rightarrow$  sophia and virtue (from [6] & [7])
  - [9] Sophia and virtue  $\Rightarrow$  PP (from [2])
  - [10] Philosophy  $\Rightarrow$  PP (from [8] & [9])
  - [11] Philosophy  $\Leftrightarrow$  PP (from [5] & [10]).
40. For the restriction of genuine or complete virtue to the philosopher and a brief introductory discussion of the issues surrounding genuine or complete virtue and demotic or imperfect virtue see Bobonich (2002, pp. 42–5). See also Irwin (1995) and Kamtekar (1998). See also *Meno* 88b1–8 and Scott (2006, pp. 146–53).
41. See 490d1–2, 492d5, 494e3–5, 498e4, and 501e4–5.
42. See also *Republic* 495c8 and Nettleship (1925, p. 203). While Socrates does appear to explicitly concede that philosophers are in truth useless and vicious (the claim at 489b3 only

- concedes their uselessness, but 487d10 and 495c8 appear to concede both uselessness and viciousness), he cannot quite mean what he says. As we will see he does literally concede the truth of the claim that philosophers as he understands them are useless in current cities, but he does not literally concede the truth of the claim that philosophers as he understands them are vicious. Rather he concedes the truth of this claim only understood in one of two ways: (1) those with the natural abilities necessary for philosophers are vicious in current cities and (2) those who imitate philosophers in current cities are vicious. The claim is false, however, understood as genuine philosophers (i.e., philosophers as he understands them) are vicious.
43. See Adam (1902, p. 9), Nettleship (1925, p. 204), and Keyt (2006). Keyt's recent essay devoted to Plato's so-called ship-of-state analogy is the most complete treatment of the image that I am aware of. I am very much in sympathy with Keyt's goal of establishing this analogy alongside the more famous analogies of the Sun, Line and Cave that immediately follow upon the conclusion of the current argument. I also found much in Keyt's interpretation of the analogy with which to agree, although my current concerns will lead me to focus on different aspects of the analogy than does Keyt.
  44. See also 489b8–9.
  45. See n. 42 above.
  46. Forms of *phusis* occur regularly throughout this section of the text. It is essential, however, to keep distinct the nature of philosophy (i.e., what philosophy is) from the natural abilities necessary for philosophy.
  47. See also *theou moiran* at 493a1–2. Indeed, the entire passage from 492e2–493a3 evidently alludes to *Meno* 99b5–100b4, as Adam (1902, p. 22) recognizes. Notice that divine dispensation plays two roles in the current passage. As in the *Meno*, it explains how genuine philosophers can arise in the current climate. But it also explains how genuine philosophers can become rulers in the current climate; see 499a10–c6.
  48. Alcibiades is often thought to be alluded to in this passage depicting the corruptive influence of family, friends, and flatterers on natural abilities (see, e.g., Adam 1902, p. 25; Nettleship 1925, p. 207; Ferrari 2000, p. 198 n. 12; Pappas 1995, p. 119; and Scott 2006, p. 167, *pace* Annas 1981, pp. 186–7), but it may be just as likely that Plato has *Meno* in mind, especially given the description of the one corrupted as “rich, wellborn, good-looking, and tall” (*Rep.* 494c6–7). See *Meno* 71b4–8: “Or do you think that it is possible for someone who does not know at all who *Meno* is to know whether he is fine or wealthy or well-born or the opposite of these?”
  49. The *Meno* also considers the poets at 95c9–96a5—however briefly—which do not get considered in the *Republic* at this point, although they certainly do earlier and later on in the *Republic*.
  50. Actually, Anytus maintains that sophistic education is harmful as well, but the argument against sophistry that appears to receive Platonic endorsement is the one presented at 95b9–c8 after Anytus leaves the conversation. Indeed, Plato's attitude with respect to sophistry is somewhat ambivalent in both dialogues. Both dialogues begin by defending sophistry against the charges leveled at it by traditionalists (*Meno* 91c6–92c7 and *Republic* 492a5–493a5) and then go on to find fault with sophistry (*Meno* 95b9–c8 and *Republic* 493a6–494a10).
  51. Scott (2006, p. 218 n. 5): “*Republic* VII 501c4–502a4 contains some striking similarities to the ending of the *Meno*. Socrates talks of the need to persuade the *demoi*, and to replace their anger with mildness.”
  52. Bostock (1986, p. 168) raises roughly this objection.
  53. For a perspicuous account of the difficulties here see Bailey (2005, pp. 97–9); see also Gentzler (1991).
  54. See, for example, Foster (1937), White (1984), and Pappas (1995, p. 55).
  55. What sort of *aitia* is at stake here is worthy of further consideration, but one suspects that it is closest to Aristotle's formal *aitia*. See, e.g., Vlastos (1969).
  56. Notice the preceding argument does not depend on the interpretation of Glaucon's challenge I described two paragraphs back. What is critical is the distinction between two kinds of

consequences—consequences “caused” by the Form alone and consequences “caused” by the Form and contingent circumstances, whether or not this distinction is at play in Glaucon’s challenge.

57. N.B. I am careful here not to suggest that Plato has made progress, nor even that Socrates’ professed ignorance in the *Meno* is genuine, in order to leave open the questions of developmentalism mentioned above. Should we ask why Plato would present us with an incomplete application of the method of hypothesis with respect to the nature of virtue, the issues surrounding developmentalism can no longer so easily be put to one side. See Weiss (2001) for an answer to this question that generally sidesteps the issues surrounding developmentalism.
58. I would like to thank Roslyn Weiss, Michelle Jenkins, and the participants at the conference in Pyrgos, Greece in the summer of 2006 and at the Arizona Colloquium in Ancient Philosophy in February 2007 for helpful comments on various versions of this essay.

## Bibliography

- Adam, J. 1902. *Plato’s Republic*. Cambridge: Cambridge University Press.
- Annas, J. 1981. *An Introduction to Plato’s Republic*. Oxford: Clarendon Press.
- Bailey, D.T.J. 2005. “Logic and Music in Plato’s *Phaedo*.” *Phronesis* 50:95–115.
- Bedu-Addo, J.D. 1984. “Recollection and the Argument from a Hypothesis in Plato’s *Meno*.” *Journal of Hellenic Studies* 104:1–14.
- Benson, H.H. 2003. “The Method of Hypothesis in the *Meno*.” In *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, edited by J.J. Cleary and G.S.J. Gurtler. Vol. 18.
- Benson, H.H. 2005. “Plato’s Rationalistic Method.” In *Blackwell Companion to Rationalism*, edited by A. Nelson. Oxford: Blackwell Publishing.
- Benson, H.H. 2006. “Plato’s Method of Dialectic.” In *Blackwell’s Companion to Plato*, edited by H.H. Benson. Oxford: Blackwell Publishing.
- Bobonich, C. 2002. *Plato’s Utopia Recast, His Later Ethics and Politics*. Oxford: Clarendon Press.
- Bostock, D. 1986. *Plato’s Phaedo*. Oxford: Clarendon Press.
- Carone, G.R. 2001. “*Akrasia* in the *Republic*: Does Plato Change His Mind?” *Oxford Studies in Ancient Philosophy* 20: 107–48.
- Cornford, F.M. 1957. *Plato and Parmenides*. Indianapolis: The Bobbs-Merrill Co.
- Dancy, R.M. 2004. *Plato’s Introduction of Forms*. Cambridge: Cambridge University Press.
- Ferrari, G.R.F., ed., and T. Griffith, trans. 2000. *Plato: The Republic*. Cambridge: Cambridge University Press.
- Foster, M.B. 1937. “A Mistake of Plato’s in the *Republic*.” *Mind* 46:386–93.
- Gallop, D., ed., trans. 1975. *Plato: Phaedo*. Oxford: Oxford University Press.
- Gentzler, J. 1991. “*Sumphonein* in Plato’s *Phaedo*.” *Phronesis* 36:265–77.
- Grube, G.M.A., trans. 2002. *Plato: Five Dialogues*. Indianapolis: Hackett Publishing Co.
- Grube, G.M.A., and C.D.C. Reeve, trans. 1992. *Plato: Republic*. Indianapolis: Hackett Publishing Co.
- Halliwel, S. 1993. *Plato: Republic 5*. Warminster, England: Aris & Philips.
- Irwin, T. 1995. *Plato’s Ethics*. Oxford: Oxford University Press.
- Kahn, C.H. 1996. *Plato and the Socratic Dialogue*. Cambridge: Cambridge University Press.
- Kamtekar, R. 1998. “Imperfect Virtue.” *Ancient Philosophy* 18:315–40.
- Kanayama, Y. 2000. “The Methodology of the Second Voyage and the Proof of the Soul’s Indestructibility in Plato’s *Phaedo*.” *Oxford Studies in Ancient Philosophy* 18:41–100.
- Keyt, D. 2006. “Plato and the Ship of State.” In *The Blackwell Guide to Plato’s Republic*, edited by G. Santas. Oxford: Blackwell Publishing.
- Nettleship, R.L. 1925. *Lectures on the Republic of Plato*. London: Macmillan and Co.
- Pappas, N. 1995. *Plato and the Republic*. New York: Routledge.
- Robinson, R. 1953. *Plato’s Earlier Dialectic* (2nd ed). Oxford: Oxford University Press.
- Scott, D. 2006. *Plato’s Meno*. Cambridge: Cambridge University Press.

- Vlastos, G. 1969. "Reason and Causes in the *Phaedo*." *The Philosophical Review* 88(3):291–325.
- Weiss, R. 2001. *Virtue in the Cave: Moral Inquiry in Plato's Meno*. Oxford: Oxford University Press.
- White, N.P. 1984. "The Classification of Goods in Plato's *Republic*." *Journal of the History of Philosophy* 22(4):393–421.

# Chapter 14

## Reasoning About Justice in Plato's *Republic*

Anthony W. Price

Gerasimos Santas has devoted his career to displaying that in order at once to clarify the content of Plato's dialogues, and to realize Plato's purpose in writing them, we must do philosophy ourselves. We have to play the roles both of handmaiden, attending to all the details of the text, and of apprentice, not leaving the philosophy to Plato but joining in the argument. Santas sets us an example of clarity of thought, lucidity of expression, and—to convey a positive quality negatively—lack of egocentricity to set beside that of Gregory Vlastos.

If Plato intends his reader to share his own risks, the constructive interpreter must take on the perils of constructive philosophy. What *his* readers can reasonably expect is not that he play safe, but that he refresh their own reading of Plato. As a sample of such enlivening, I here offer a response to Santas (2006). This invites our attention as bringing equally to life points in the text that do fit it, and points that appear not to.

The paper effectively expands a pregnant paragraph of his earlier and excellent book, *Goodness and Justice*. To quote from that paragraph is to summarize the present paper:

Polemarchus relies on the wisdom of the poets, the moral educators of the Greeks. Thrasymachus proceeds as an empirical political scientist: on the assumption that the justice of any city is to be found in (or is identical with) its system of laws, he defines justice on the basis of what he thinks is a general observable fact about all such systems, that they are created to promote the interest of their rulers. Glaucon uses a contractarian methodology: he assumes certain facts about humans (they are rational creatures) and their environment in nature (moderate scarcity of the things they want and consequent conflicts among them), and relates how such creatures would reason to a system of laws: each gives up equally the freedom to harm others in return for the equal security of not being harmed by them . . . The functional methodology Plato uses is different from all these. The *Republic* discusses not only different systems of justice but also suggests different methodologies for discovering and justifying different systems. And methodologies make a difference to what results are reached. (2001, p. 76)

---

A.W. Price (✉)  
Birkbeck College, London, UK  
e-mail: a.price@bbk.ac.uk



This is a novel and elegant structuring, one that is more plausible than most original readings, and more original than most plausible ones. Its later, ampler presentation adds valuably to the book in its fuller treatment of Glaucon's contractarian narrative. Yet what I want to discuss is first the general framework, and then "the functional method" itself. Reflecting upon the second takes us to the heart of *Republic IV*.

## 14.1 Alternative Methodologies

It at first appears that the theories are creatures of the methodologies: "The significance of methods of reasoning about justice become evident: different methods might give us different results, as they actually do in the *Republic*" (Santas 2006, p. 125). However, Santas concludes by conceding that "the different results were obtained by using different substantive assumptions as well as different methods" (pp. 144–5); so it would seem that wrong results might come of sound methods but false assumptions. Thus we have to assess the different roles of method and assumptions in each case.

When Polemarchus breaks in to reaffirm the definition "if we are to believe Simonides" (*Rep.* I 331d4–5, Griffith transl.), is this, as Santas affirms, "simply an appeal to authority," perhaps even "divine" (2006, p. 126)? When Socrates asks him to clarify what *he* says Simonides is right to say about justice, he offers the formula "It is just to pay everyone what is owed to him," remarking "That's what he says, and I think he's right" (331e1–4). If he appeals for support to Simonides, he also offers him *his* support. When Socrates ironically ascribes an unwelcome implication to "you and Homer and Simonides" (334b3–4), this does not have the absurdity of "you and God." No doubt Polemarchus lacks the independence of mind that enables Thrasymachus to cite the adage that justice is another's good only to gloss this as the advantage of the stronger (343c3–4, on which see Shorey 1930, *ad loc.*, citing evidence from, among others, Aristotle, *NE* V.6 1134b5–6); also that which makes it entirely incidental when Socrates later appeals to what "we have often heard others say," as well as what "we have often said ourselves," for the view that "doing one's own job, and not trying to do other people's jobs for them, is justice" (IV 433a8–b1). A taste for citation can have many grounds; the fact of citation proves nothing.

Santas views Thrasymachus as anticipating Aristotle's empirical investigation of constitutions. He says that Plato "explicitly portrays" him as "employing a certain method for arriving at his account of justice" (2006, p. 126), a method exemplified by the "big empirical generalization" that "in each form of government the ruling party enacts laws to its own advantage" (p. 127). He even identifies it as "another generalization" that "each form of government proclaims that justice consists in observing the laws it has enacted" (*ibid.*). This may be implicit at I 338e5–6 (it punishes men "on the ground of breaking the law and acting unjustly"). Yet Socrates takes Thrasymachus himself to be saying that it is just to obey the laws (339b7–8, 340a7–8), and Polemarchus supposes him to imply that, when obeying

the laws and benefiting the rulers come apart, it is just to obey the laws (340a4–6). Thrasymachus' reply that rulers do not err *qua* rulers in prescribing their own advantage probably has the corollary that laws *stricto sensu* are the enactments of rulers *qua* rulers (cf. 340e2–341a2); which should always reconcile obeying the law with benefiting the rulers.

Santas takes Socrates' second argument (341c4–342e11) to confirm that “the method is an empirical investigation and the main premise an empirical generalization” (2006, p. 128); yet the argument proceeds in accordance with “the most precise account” (*ho akribestatos logos*, 341b8), which makes use of the *qua*-construction rashly licensed by Thrasymachus himself, and so is not empirical. Santas calls it an “assumption” that it is just to obey the law, and a “definition” of justice that it is the advantage of the ruler (2006, p. 127). One way of understanding Thrasymachus is to ascribe to him a *verbal* definition of justice as obeying the law, and a *real* definition of justice as thereby benefiting the rulers. (It is true that 344a3–c8 count the tyrant as the paradigm of injustice; but we may suppose that this is the *nouveau tyran*, who is acting against the surviving laws and constitution of the democracy that he has overthrown.) The real definition states the real nature of justice, which is fixed by its actual function; what this is becomes evident “if one reasons rightly” (339a2–3), and so clearly isn't axiomatic.

What has misled Thrasymachus? It can hardly be empiricism as such. Indeed, Socrates' refutation veers opportunistically between the “precision” of the *qua*-construction (note the emphatic *alêthôs* and *akribôs* of 345c1–3, 346b3, d2) that is required to distinguish one art from another (346a1–d12), and the impressionism of a generalization (that the good are reluctant to rule, 347b5–d8) whose truth cannot be purely conceptual. Moreover, his own later investigations rest no less on empirical data, as Santas recognizes: “The substantive conclusions Plato reaches in the *Republic*, in ethics and political philosophy, are the result not only of the functional theory but also of many empirical assumptions he makes about human beings and their cities” (2001, pp. 77–8). Thus Socrates concedes a crucial element of contingency when he remarks, “The origin of a city lies, I think, in the actual fact that each of us is not self-sufficient (*tunchanei hêmôn hekastos ouk autarkês*); we have all sorts of needs” (II 369b6). When he adds that things go better when “one person does a single task which is suited to his nature” (370c4–5), he must be generalizing from human experience. (Compare much that is said, for example about canine character, between 375a12 and 376a10.)

What Santas finally finds crucial in Thrasymachus' position is “the assumption that makes the empirical investigation of justice and its result possible: that justice is to be found in the laws of each society or form of government” (2006, p. 142). As I have noted, the assumption itself is rather semantic than contingent. It may be needed to make *certain* empirical claims (viz. those that he makes about the motivations of legislators) immediately relevant; it is not needed for *any* empirical claims (even those of Socrates). Strikingly, it is taken over by Glaucon as a stipulation within the social contract: “They [the contractors] call lawful and just that which is laid down by the law” (359a3–4). If Thrasymachus' “method is wrong-headed” (Santas 2006, p. 142), this is not because of its empiricism.

You might ask: “Does it matter for Santas if method and theory are really one?” Well, “different methods, different theories” sounds interesting; “different themes, different elaborations” risks tautology.

Santas’ account of Glaucon is not careful to demarcate method from result. We meet the heading “The Contractarian Method of Glaucon,” but then read of “Plato’s contractarian account of what justice is” (p. 129). We are alerted to the workings of no methodology distinguishable from a resultant theory.

To supplement the omission, one might quote the wording that frames Glaucon’s initial summary statement of his contractarian hypothesis, introducing and concluding it: “Now listen to the first thing I said I was going to talk about—what sort of thing justice is, and how it arises . . . That is what it [justice] is like, and those are the kinds of causes which gave rise to it, according to this theory” (358e1–2, 359b4–5, cf. 359a4–5). So one might take the method simply to be this: look for the origin. Yet it would then confuse any application of Santas’ contrasts that Socrates adopts the same approach: “Suppose that we were to examine the origin of a city discursively, would we also see the origin in it of justice and injustice?” (369a5–7). This is how he enters upon a narrative of a sequence of specializations within an emergent city.

The relation of method to theory is not clarified when Santas writes, “Plato does not explicitly or implicitly criticize Glaucon’s method for discovering the nature of justice by looking for a contractarian origin of justice” (2006, p. 142); for surely Glaucon’s *account* of justice is contractarian. Is Santas connecting a method to a theory, or different theories, when he continues, “Nor does he [Plato] consider the question whether his own conception of justice would be reached by using Glaucon’s method” (*ibid.*)? As I noted, Glaucon plainly supposes what Thrasymachus apparently presumes, that the just is the lawful (compare I 340a7–8 with II 359a3–4). Thrasymachus then denies a subject any reason to be just on the ground that laws actually serve the interests of legislators; Glaucon finds a reason to be just in men’s actual inability to be sure, when both are permitted, of *inflicting* without *suffering* wrong. What is the appeal of the social contract? It is one that it shares with Socrates’ story of how specialisms multiply in response to needs: “Let us create a city discursively from the beginning. It is the product, apparently, of our needs” (369c9–10). Both narratives confirm the desirability of the developments they imagine by attributing them to individual initiative and consent, and in this way both contrast with Thrasymachus’ explanations, which debunk law-abidingness as the dupe of an exploitative ideology. This raises a different question, not of how to reach a theory, but how to justify one: could Plato have grounded his own system of justice upon contract theory?

Santas is not sanguine. On the one hand, “Conceivably, Plato thought that if we supply what he thought were true or reasonable assumptions, men in a state of nature would choose his principle of social justice over the state of nature” (2006, p. 142). On the other, “Plato’s own theory of human good, with its radical downgrading of freedom and the goods and pleasures of ordinary men . . . seems hardly something that we can attribute to men in a state of nature” (p. 143). Socrates has not faced quite this concern when he concludes a famous section of the *Republic* as follows:

“Our arrangements are the best, if only they could be put into effect, and while it is difficult for them to be put into effect, it is not impossible” (VI 502c5–7). He may require the emergence, within a long transition from a state of nature towards the free realization of his utopia, of civic arrangements that, by their political failures and educational successes, pave the way for a popular appreciation of the very dialogue of which he is the protagonist.

However, we should not allow such speculations to obscure a contrast. It is not in order to confirm that his utopia is not dystopian that Socrates feels the need to argue that it is not necessarily impracticable. His concern, on his own and his interlocutors' behalf, is that they not “be a laughing-stock” for advancing proposals that are “just wishful thinking” (499c4–5). And yet, however apprehensively, it is “compelled by the truth” (b4–5) that they have been saying what has been said. Within Plato's utopianism, the possibility of consent at once popular and rational is a pious hope and not a criterion of truth. Within Glaucon's contractarianism, it is what establishes that justice is a second-best virtue: that rational agents could freely consent to establish laws indicates that insecure subjects have reason to obey them. This, of course, is why social contract theory belongs within liberal and not paternalist political theory.

Is this a methodological difference between alternative justifications of justice? Perhaps it is. It is certainly a gulf between political philosophies.

## 14.2 The Functional Method

When we turn to what Santas calls “the functional method of Plato,” it is again unclear what counts as method, and what as theory. He asks, “Can we discover a method from what he actually does between his starting point and his definitions?” and begins a reply by remarking, “Well, Plato has Socrates begin by dividing the question into two, what is a just city and what is a just person, and starts with a just city” (2006, p. 133). Here is certainly a method, which we might call *macroscopic*, citing a familiar passage: “Maybe justice will be on a larger scale in what is larger, and easier to find out about. So, if you approve, why don't we start by finding out what sort of thing it is in cities? After that, we can make a similar inquiry into the individual, trying to find the likeness of the larger version in the form the smaller takes” (II 368e7–369a3). Of course, this rests on the presumption of what Santas calls “isomorphism”: Socrates supposes that the ascription of justice to cities and individuals indicates that they can share a single quality. As he will later reason, “If you have two things—one larger, one smaller—and you call them by the same name, are they like or unlike in respect of that which gives them the same name? . . . So the just man in his turn, simply in respect of his justice, will be no different from a just city” (IV 435a5–b2). Behind the methodology lies a theoretical assumption.

As is familiar from Williams (2006), the macrocosm-microcosm model is only one of two that Plato employs. The other is of group-member dependency. Any quality of a city derives from the citizens who possess it (435e1–436a7), and from their displaying it within the city: thus Guardians make it wise in exercising their

wisdom on behalf of the city as a whole (428c11–e1), while auxiliaries make it brave in exercising their courage on its behalf (429b1–6). The first model is one of *isomorphism*, the second of *derivation*. The conjunction of the two risks incoherence. According to the derivation model the justice of a citizen is external, but according to the isomorphism model it is internal: we read explicitly that the justice of an individual consists in his doing his own business not externally, but within his soul and in respect of its parts (443c9–d2). So a just city is one whose citizens are just in exercising justice *within it*; yet just citizens are those who are just in enjoying justice *within themselves*. However, at least the broad lines of a resolution are clear: internal and external justice must be distinguishable aspects of one and the same disposition of soul. (So, doubtless among others, Price 1997, sect. 3.)

Both models can be taken to support a city-centered approach towards the definition of justice: since justice is isomorphic, we may start with the plainest case (justice of city), and apply its lessons to the case that most concerns us (justice of soul); since it involves a relation of derivation, we may identify the justice of citizens with whatever quality in them makes a city just (which *focuses* the justice of a citizen upon the justice of a city). Perspicuously just is the city of which it can be said that “each of the three types of nature in it was performing its own function” (435b4–5). This is a city that owes its justice to the observance by *every* member of it of the principle that each person do his own job (433d1–4). We can then also count as just a soul that possesses the same three elements in the same condition (435b8–c2). What makes Socrates conscious of his present approach as one among others is that he does not overrate it: “I have to tell you, Glaucon, that in my view we are certainly not going to find a precise answer to our enquiry by the kind of methods (*methodoi*) we are using at the moment in our argument. There *is* a way of getting there, but it is longer and more time-consuming. Still, we may be able to get an answer which is no worse than our earlier answers and investigations” (c9–d5). The context indicates that it is the macroscopic method that he has in mind.

That Socrates would count use of the notion of a function as a distinctive method is made less likely by its ubiquity. Within the *Republic*, it serves first against Polemarchus: “So, Polemarchus, it is not the function of the just man to treat his friend or anyone else badly. It is the function of his opposite, the unjust man” (I 335d11–12). It is then deployed against Thrasymachus: “Each skill performs its own function, and benefits the object of which it is the skill” (346d5–6); “It is the function of injustice to produce hatred wherever it goes” (351d9). It is in that context that Socrates spells out two features of the notion that Santas is right to stress: the function of a thing is that which it alone brings about, or brings about better than anything else (353a10–11); and for anything with a function there is an excellence that enables it to perform its function well (c6–7). Socrates evidently views such ideas as commonplaces that his interlocutors, whatever *their* methods of investigation, can only play along with. The notion of a function, and of excellences whereby things function well, is for Plato not an alternative method (or, indeed, theory), but an *idée fixe*.

As Santas instances, Socrates makes explicit use of the notion when describing justice in a city. Thus he asks, “Shouldn’t we be persuading them [the

Guardians]—and everyone else likewise—to be the best possible practitioners of their own particular function?” (IV 421c1–3; cf. III 406c3–5). Santas concedes that this is not explicitly applied to justice in the soul, but infers from the isomorphism that the same structure, of three parts associated with three functions, must still be in play (2006, pp. 138–9). What we have now to consider is how well the notion serves Plato's purposes.

Santas spells out the natural capacities of each class as follows (2001, p. 87): “Those best suited to govern are people of inborn high intelligence and appropriate education; those best suited to defend the city are persons of inborn high spirit and appropriate education; and those best suited to provision the city are persons of inborn abilities and education for arts and trades.” If we adopt his terminology to capture Plato's distinction, this defines functions that are *optimal* rather than *exclusive*. Which is unsurprising: every member of the city is a human being, with a complete human soul each of whose parts is associated with one of the three capacities. However, problems arise once we descend from Guardians proper (whom I shall call “philosophers”) to auxiliaries and artisans. For instance, are philosophers less able “to defend the city” than auxiliaries are? How can they be, when the education of philosophers contains a military phase that they share with auxiliaries? When Socrates starts to describe the nature needed by his Guardians, he does not distinguish the two classes (II 374d8–376c5). Instead, he relates gentleness to friends and hostility to strangers to a love of what one knows, and concludes that a Guardian must be “by nature a lover of wisdom, spirited, swift and strong” (376c4–5). It appears that, when philosophers are later distinguished from auxiliaries, this is because the first have an ability that the second lack, and not vice versa. But this produces a problem: as Santas has written, “If there are things which can be good but have no function, the functional theory cannot account for their goodness” (2001, p. 69). For it seems that the nature of auxiliaries supplies them with neither an exclusive nor an optimal function.

It is true that the idea of an optimal function can be played two ways. (See Irwin 1977, p. 333 n. 34, p. 343 n. 28; Santas 2001, pp. 78–81.) It is clear from the context of I 353a10–11 that what something “brings about better than anything else” is primarily what it brings about well, and other things bring about worse. (Thus a pruning-knife is better at pruning than a carving-knife; a1–5). However, Socrates surely also has in mind what something brings about better than it brings about other things. Otherwise, we risk a divergence between what is best for the city and best for the individual: if we do link function and happiness, it should best serve the happiness of the city that I do (among the things that need to be done) what I am better than others at doing; yet it presumably best serves my own happiness that I do what I do better than other things. (And more can go wrong: there may be essential tasks to be performed within the city which too few are particularly good at in either sense.) So we might defend Socrates from the objection above by proposing that auxiliaries should defend the city not because they are better soldiers than philosophers are, but because they are better soldiers than they are philosophers. This would be an amendment, but one that is not alien to Plato's purposes.



Parts of the soul import their own problems. Santas rests Plato's partition upon distinctive functions as much as upon his explicit criterion of conflict: "I believe that these two ideas, of exclusive functions and characteristic objects, are the main principles at work for individuating the parts of the soul; the partition itself, as based on psychic conflict and the principle of contrariety, is not sufficient for understanding the nature of each part" (2001, p. 122). Broadly, this must be right: a recalcitrant contrariety of desires evidences a conflict between parts, but this is only explanatory if the parts are sources of desires of different kinds that enjoy unequal plasticity or adaptability. (Thus, if a desire to eat arises from hunger, we can easily understand why it does not evaporate in the presence of a rational prohibition on eating.) But what category of function is in play? Santas finds it implicit in the isomorphism of city and soul that "the human soul comes with a natural division of parts . . . and psychic labors (functions) unique (exclusive) to each part" (2006, p. 139). Exclusive functions are apparently assigned to reason and appetite in this passage: "The part of the soul with which we think rationally we can call the rational element. The part with which we feel sexual desire, hunger, thirst, and the turmoil of the other desires can be called the irrational and desiring element" (IV 439d5–8). And we read much later that "measuring, counting and weighing . . . are the function of the rational element in the soul" (X 602d6–e2), presumably its exclusive function. However, we then risk an unexpected implication: if soul-parts are defined by exclusive functions, a part has only to be operative for it to be doing its own job; an active reason or appetite is ipso facto a just reason or appetite. An unjust soul becomes a conceptual impossibility.

Doubtless this is why, in defining justice, Santas shifts his attention from exclusive to optimal functions: "A soul is just when each part of it is performing that psychic function . . . which it can do best (i.e., optimally): for example, reason must rule because it can do that better than spirit can rule . . . , better than appetite can rule, and so on" (2006, p. 139). This fits better, for spirit or appetite *can* rule, i.e. dominate, as they do within corrupt souls. Perhaps parts are first distinguished from one another by their exclusive functions, and then characterized as just or unjust in respect of their optimal functions. Indeed, Santas (2001, p. 129) states this clearly: "The argument is based on the exclusive functions of the three psychic powers, as conceived in the partition argument, *and* what is required for assigning optimally to the various parts of the soul the psychic functions of ruling, defending, and provisioning." (I would prefer to speak of the psychic functions that *subserve* defending and provisioning.) It is because only reason can calculate and compare alternative courses of action that ruling the soul is the optimal function of reason, and not of appetite or spirit. Talk of "operating justly" ceases to be pleonastic.

Ascribing exclusive or optimal functions to soul-parts becomes a delicate task. Santas infers from the description of appetite as "arational" or "non-calculating" (*alogistos*, IV 439d7) that it is "not that both appetite and reason can calculate but reason can do it better, but that appetite cannot or at least does not calculate at all" (2001, p. 123). He takes it that, in order "to discover the nature of appetite itself apart from the contributions cognitive learning makes," Plato "abstracts from, mentally takes away from, appetite, other psychic activities that attach to it, notably all



learning and cognition about the objects of our appetites" (ibid.). Appetite becomes, itself, the home only of undifferentiating desires for food, drink and sex—generic objects that definitionally distinguish hunger, thirst, and sexual desire. More sophisticated desires arise from interactions between soul-parts, "combinations which generate mixed or derivative psychic states, as a result of experience, habituation, and education" (ibid., p. 124).

We may welcome this recognition of the emergence of what Freud would call *compromise formations*, here desires or pleasures that emerge from a cooperation between parts. This may well be involved in the varied enthusiasms of Plato's democratic man, whose "variegated life" is "full of all sorts of characteristics" (VIII 561e3–4), alternating unpredictably between wine and water, sloth and exercise, war and business, politics and philosophy (c7–d5). Take his philosophy: this may not be less intelligent than that of the genuine lover of wisdom; but it takes on the hedonism natural to appetite, being pursued only for fun, and not (which is harder work) in pursuit of truth. (An Anglo-Saxon here may think of certain contemporary French celebrities.) It is thus plausibly ascribable to a new formation within the soul which straddles intelligence and appetite.

Yet are we to suppose such complications whenever a mouth waters not just at the prospect of food but at the sight of an éclair, and sexual arousal is provoked not indiscriminately but selectively? (There is no suggestion of this in the *Phaedrus*.) What also of the derivative desire for wealth? Santas assigns this neither to reason nor to appetite alone, placing it rather in the category of "mixed or combined psychic activities" (2001, pp. 163–4 n. 37; cf. p. 124). Yet contrast *Republic* IX: "We called it [the third element] desiring—because of the strength of its desires for food, drink, sex and everything that goes with these—and money-loving, because money is the principal means of satisfying these desires" (580e2–581a1; cf. VIII 553c5)? Elsewhere Plato calls the appetite which reason and spirit are to watch over "the most insatiable where material goods are concerned" (*chrêmatôn phusei aplêstotaton*, IV 442a6–7; cf. VIII 549b2). Doesn't this imply that love of wealth is a property, indeed a characteristic one, of appetite?

And what are we say, for example, of a case where reason opposes a desire for money? Are we to posit a *fourth* part and faction, offspring of reason and appetite? To avoid proliferation, we may need to concede to appetite an ability, or tendency, to slide from desiring an end to desiring a means. It confirms the ascription of this to appetite that it can involve blatant illogicality: if "the pleasure which comes from money" (IX 581d5–6) is to be single, it must be the miser's rather than the spendthrift's; and the miser commits the fallacy of valuing an instrumental good as if it were an intrinsic one. *His* transition of thought is irrational and regrettable. Other reasoning that speaks to appetite may be requisite and reasonable. It is unhelpful that Socrates fails to spell out the *logos* by which reason can "persuade" appetite to fall in line (IX 554d2); for it is "agreement" between the parts that reason must rule that defines temperance within the soul (IV 442c10–d1). Though the issue is debated, I still incline to suppose (as in Price 1995, pp. 59–65) that, for better and worse, appetite has something of a head for simple inferences, its own or another's, and thereby comes to desire new things appetitively. If this is right, it has implications

for the exclusive function of reason. This would have to be reasoning of a different style that may still, by rational standards, be good or poor; if so, it remains unfinished business to demarcate the kind of reasoning that is exclusive to reason as a faction within the soul.

Such reflections are pertinent, and require pertinacity. May Santas' paper and book together prompt us to persevere.

## Bibliography

- Griffith, T., trans. 2000. *Plato: The Republic*. Cambridge: Cambridge University Press.
- Irwin, T.H. 1977. *Plato's Moral Theory*. Oxford: Clarendon Press.
- Price, A.W. 1995. *Mental Conflict*. London: Routledge.
- Price, A.W. 1997. "Plato: Ethics and Politics." In *From the Beginning to Plato*, edited by C.C.W. Taylor, 394–424. Oxford: Routledge.
- Santas, G. 2001. *Goodness and Justice: Plato, Aristotle, and the Moderns*. Oxford: Blackwell.
- Santas, G. 2006. "Methods of Reasoning about Justice in Plato's *Republic*." In *The Blackwell Guide to Plato's Republic*, edited by G. Santas, 125–45. Oxford: Blackwell.
- Shorey, P., ed., trans. 1930. *Plato: The Republic* (2 vols). Cambridge, MA: Harvard University Press.
- Williams, B. 2006. "The Analogy of City and Soul in Plato's *Republic*." In *The Sense of the Past: Essays in the History of Philosophy*, edited by B. Williams, 108–17. Princeton: Princeton University Press.

# Chapter 15

## Plato on Justice

David Keyt

### 15.1 Introduction

Justice is one of the most ubiquitous topics in Plato's dialogues, second in importance only to reason. It is discussed to some degree in almost every major dialogue including even the *Parmenides* (130b7–9, 130e5–131a2, 135c8–d1) and the *Timaeus* (41c6–8, 42b21–2), but it is only in the *Republic* that the concept is defined and the definition argued for. Consequently, any account of Plato's theory of justice must concentrate on that dialogue.

The search for a definition of "justice" is part of the larger project of the *Republic* to respond to the challenge of Glaucon and Adeimantus. Speaking as devil's advocate Glaucon classes justice among the goods chosen, not for their own sake, but for the things that come from them. People, he claims, want no shackle on their natural desire for more and more of everything, and only agree to act justly towards others in order to avoid being treated unjustly themselves. That no one is just willingly is shown, he says, by the story of Gyges' ring, a ring that bestows invisibility upon its possessor: no one who possessed such a ring could resist the temptation to become "like a god among men" by using it to satisfy his natural desires unrestrained by justice (*Rep.* II 357a1–360d7). Socrates sets out to show, contrary to this impressive challenge, that justice is good both in itself and for what comes from it and that injustice, even if it goes undetected, is injurious to the unjust person. The first step in meeting the challenge, a large one, is to understand what justice is. Only when we understand this, Socrates reasonably claims (*Rep.* I 354c1–3), will we be able to determine whether justice is good in itself or good only because of what comes from

---

This chapter originally appeared as a chapter in *A Companion to Plato* edited by Hugh H. Benson (2006; Oxford: Blackwell) and is republished by permission of Blackwell. The fact that one of its principal interpretative ideas is due to Jerry Santas makes it, I hope, an appropriate contribution to this festschrift in his honor. I am grateful to Benson, and my wife, Christine Keyt, as well as to Jerry, for helpful comments on earlier drafts.

D. Keyt (✉)  
University of Washington, Seattle, WA, USA  
e-mail: keyt@u.washington.edu

it. This paper is devoted entirely to this first step, Plato's definition of justice, and does not discuss the sort of good it is or the link Plato endeavors to forge between justice and happiness.

The burden on anyone expounding Plato's theory of justice is to fill the numerous gaps in his argument, to supply the missing premises. The further afield an interpreter must go to find appropriate premises the less credible their attribution to Plato will be. In this paper I never look beyond Plato's dialogues themselves, and only at one or two crucial junctures beyond the *Republic* itself. I never appeal to other ancient Greek philosophers or to the philosophical imagination itself. This does not mean that mine is the only way, or the best way, to fill out Plato's argument. There are alternatives (for example, Dahl 1991). (For the interpretative strategy followed in this paper and its ramifications, see Cohen and Keyt 1992.)

I assume that Socrates is Plato's spokesman in the *Republic* and that the Eleatic and Athenian Strangers speak for Plato in the *Statesman* and the *Laws* respectively.

## 15.2 *Phusis* and *Nomos*

In the *Laws* the Athenian Stranger considers two connected ideas about justice advanced by some poets and prose-writers identified only as certain "modern wise men" (*Laws* X 886d2–3). These ideas are that justice is an unstable artifact of human contrivance and that might makes right. According to the modern wise men:

[T]he just things are not at all by nature (*phusei*) but people are continually disputing with one another about them and are forever changing them, and whatever changes they make at any time are each at that time authoritative, having come into existence by art (*technê(i)*) and by the laws (*nomois*) but not in any way by nature. All these things. . . are the theme of men considered wise by young people—prose-writers and poets—who maintain that what is most just is what a person can win by force (*Laws* X 889e6–890a5).

The claim of the modern wise men that the just things exist by art (*technê (i)*) and by the laws (*nomois*) but not in any way by nature (*phusei*) exploits a favorite antithesis of 5th and 4th century Greek philosophy, that between *nomos* (law or convention) and *phusis* (nature) (for which see *Prt.* 337c6–d3 and *Grg.* 482e2–484c3, 488d5–489b6). In this antithesis *nomos* is associated with artificiality, diversity, and variability, *phusis* with truth, sameness, and invariability (see in particular Aristotle, *SE* 12 173a7–18 and *NE* I.3 1094b14–16). To claim that the distinction between what is just and what is unjust exists by *nomos* only and not in any way by *phusis* is to claim that it has no firmer basis in reality than that between, say, Greek and barbarian (for which see *Pol.* 262c10–d6).

This claim leads in Plato's view to Protagorean moral relativism. If the just is simply the lawful and laws are always being changed, then what is just is relative not only to each polis, but to each point in time in each polis. In the *Theaetetus*

Socrates imagines Protagoras saying that “whatever things *seem* just and fine to each polis *are* so for it as long as it holds by them” (167c4–5; see also 172a1–5) and claims that “with regard to things just and unjust, pious and impious [the followers of Protagoras] are ready to insist that none of them has by nature (*phusei*) a being (*ousian*) of its own, but rather that what seems to people in common to be so is true, at the time when it seems so and for as long as it seems so” (172b2–6).

As the Athenian Stranger indicates, the doctrine that the just is the lawful carries in its train the unsavory doctrine that might makes right. To connect the two all that is needed is the plausible assumption that a polis’s laws are in the hands of the stronger. If (1) the just in a polis is what is lawful in it and if (2) those who make and enforce a polis’s laws—the polis’s rulers—are those who monopolize the coercive force in the polis, then, as the modern wise men claim, (3) what is just and what is won by force are the same. This conditional, it is worth noting, expresses a major part of Thrasymachus’ argument that the just is the advantage of the stronger (*Rep.* I 338d7–339a4). The second conjunct of its antecedent, proposition (2), is difficult to deny since this comes close to being a definition of a ruler; thus, anyone who has a reason for rejecting its consequent, proposition (3), has a reason for rejecting the first conjunct of its antecedent, proposition (1). But anyone who thinks that the forced is not *ipso facto* just has just such a reason.

If the just is not the same as the lawful—if “just law” is not a pleonasm nor “unjust law” an oxymoron—we need a standard of justice beyond law. Plato finds this standard, of course, in his realm of Forms. In the *Parmenides* Socrates is confident that there is a Form of Justice “itself by itself” whether or not there are Forms of such things as man, fire, water, hair, mud, and dirt (*Prm.* 130b7–d9); in the great myth in the *Phaedrus* the discarnate soul beholds Justice itself in the place beyond heaven (*Phdr.* 247c3–d6); and in the *Republic* the Form of Justice is one of Socrates’ first examples of a Form (*Rep.* 476a4–5) and the only Form, aside from the Form of the Good, mentioned specifically in the Allegory of the Cave (*Rep.* 517e1–2). Plato envisages the Forms as incorporeal entities (*Phd.* 65d4–66a10; *Sph.* 246b8) without color, shape, or solidity (*Phdr.* 247c6–7) existing beyond time and space (*Ti.* 37c6–38c3, 51e6–52b2). Having the features of truth, sameness, and invariability (*Rep.* V 476a4–7, 478e7–479a5, 479e7–8), they fall on the *phusis* side of the *phusis-nomos* antithesis. Plato is thus able to refer to the Form of Justice as “the just by nature” (*to phusei dikaion: Rep.* VI 501b2) and, in general, to identify his world of Forms with the realm of nature (*Phd.* 103b5; *Rep.* X 597b5–7, c2, 598a1–3; *Prm.* 132d2). In so doing he provides an avenue—for anyone who can countenance Forms—for an appeal beyond law to nature. Some laws will be just by nature, and some will not. (For just and unjust laws see *Laws* IV 715b2–6 and VII 807c4; and for Plato’s concept of nature see Morrow 1948.)

The issue shifts now to the content of the Form of Justice. What does one who apprehends the Form of justice apprehend? Plato begins his complex answer to this question by analyzing the justice of a polis—political justice.

### 15.3 Political Justice

Plato's ultimate goal in Books II through IV of the *Republic* is to discover what justice is in an individual soul, or psyche (*psuchê*), rather than what it is in a city, or polis. The definition of political justice is ostensibly only a way-station on the road to the definition of psychic justice, though in the overall structure of the dialogue the way-station threatens to overshadow the ultimate terminal. Socrates claims that the justice in a polis should be the easier to apprehend because, a polis being larger than an individual, justice should be more prominent (*pleiôn*) in it (*Rep.* II 368e). Plato cannot mean that political justice is easier *to perceive* than psychic justice, as large letters are easier to see than small; for justice, unlike beauty, is not a sensible property (*Phdr.* 250b1–e1). What he must mean is that political justice is easier *to comprehend* than psychic. This does turn out to be the case for justice as Plato conceives it; for the tripartite division of a polis that underlies his definition of political justice is much easier to understand than the corresponding division of the psyche underlying his definition of psychic justice (*Rep.* IV 435b4–d9).

In searching for a definition of political justice Plato focuses on what Socrates describes as “the beautiful polis” (*hê kallipolis*, *Rep.* VII 527c2), which we shall call “Kallipolis.” Socrates hopes to find justice in Kallipolis because Kallipolis is completely good (*teleôs agathê*): “I think our polis, if indeed it has been correctly founded, is completely good. . . Clearly, then, it is wise, brave, temperate, and just” (*Rep.* IV 427e6–11). Wisdom, bravery, temperance, and justice are virtues, or excellences (*aretai*). Socrates is inferring that Kallipolis has certain *aretai* because it is *agathê*. Verbally the step is a small one, from adjective (*agathê*) to corresponding noun (*aretê*): Kallipolis, being “excellent,” has certain “excellences.” Philosophically the step is larger and more problematic; for Socrates does not say how the virtues, or excellences, of a polis—its wisdom, bravery, temperance, and justice—are connected to its goodness.

But he does provide a clue in the function argument at *Republic* I 352d8–354a12. The importance of this passage for understanding the larger argument of *Republic* II–IV has been demonstrated by Jerry Santas (1985; 2001). The function argument in the passage cited is one application of what Santas calls “the functional theory of the good.” This theory consists of three definitions. (1) The *function*, or *ergon*, of each thing that has a function is (a) what it alone can do or (b) what it can do better than anything else (*Rep.* I 353a10–11). The function of the eyes is to see; the function of a knife is to cut. Santas calls a function of type (a) an “exclusive” function and a function of type (b) an “optimal” one. Eyes are defined by their exclusive function since an animal can see with no other organ; a pruning knife, on the other hand, is defined by its optimal function since other kinds of knives can be used, though not so efficiently, for pruning. (2) A thing that has a function is *good*, or *agathos*, if it performs its function well (see *Rep.* I 353e4–5). Good eyes see well; good knives cut well. (3) The *virtue*, or *aretê*, of anything that has a function is that by means of which it performs its function well (*Rep.* I 353c5–7; see also X 601d4–6). The virtues of a knife blade are the qualities that enable it to cut well, such as sharpness and hardness. The functional theory of the good allows us, then, to bridge the gap

between a city's goodness and its virtues. To apply the theory we need to answer three questions: (1) What is the function of a polis? (2) What is it for a polis to function well? (3) What qualities allow it to function well?

The first step, then, is to determine the function of a polis. What is it that only a polis can do or can do better than anything else? Plato never addresses this question directly. This no doubt is what gives rise to the general disagreement among scholars about what the goodness of Plato's ideal polis consists in. Julia Annas attributes its goodness to its organization (Annas 1981, p. 110); David Reeve claims that it is completely good because its citizens are maximally happy (Reeve 1988, p. 84); and Nicholas White, vehemently denying that either the happiness or the goodness of its citizens has anything to do with the matter, finds its goodness in its cohesiveness and resistance to destruction (White 1979, pp. 39, 114).

We may be able to resolve this dispute by going beyond the *Republic* and considering some of Plato's remarks about statesmanship (*politikê technê*). We find a list of the functions of statesmanship in the *Euthydemus*:

Then the other functions (*erga*), which someone might say belong to statesmanship—these perhaps would be many, such as making the citizens wealthy and free and without faction—all these appeared neither bad nor good; but [statesmanship] had to make [the citizens] wise and give them a share of knowledge, if this was to be the [art] that benefited them and made them happy. (*Euthd.* 292b4–c1)

According to this passage the function of a statesman is to promote the wealth, freedom, domestic tranquility, wisdom, and happiness of the citizens of his city, and to save them from poverty, slavery, faction (*stasis*), folly, and wretchedness. It seems plausible to assume that the function of a true statesman (*politikos*) is to foster the well-functioning of his state (*polis*). Combining this idea with Socrates' claim that poleis are created by human needs (*chreia*) (*Rep.* II 369c9–10), it would seem to follow that the function of a polis is to meet the needs of its citizens for the five specified goods and to save them from the five corresponding evils.

That we are on the right track—that this is, indeed, Plato's implicit notion of the function of a polis—is borne out by his description of his ideal polis; for the social, economic, and political institutions of Kallipolis address exactly these needs. Wise leadership is provided by its rulers; the city's freedom is protected by its warriors; and the need for food, shelter, and other material goods is met by its workers. The community of wives and children and the absence of private property among the rulers and warriors are designed to prevent faction, or *stasis* (*Rep.* IV 422e3–423d6, 461e5–465c7). The one need that is not addressed directly is the need for happiness. But Plato does not seem to conceive of happiness as a distinct good over and above the other goods but rather as a natural product of them. One of Socrates' remarks can, at any rate, be so interpreted. In response to the objection that in depriving the warriors and rulers of Kallipolis of gold and silver and private property he is also depriving them of happiness, Socrates replies that they along with the other citizens must be persuaded and compelled to "be the best possible craftsmen at their own work" and that "in this way, as our whole polis grows and is well governed,



one must let nature (*hê phusis*) allot each group its share of happiness” (*Rep.* IV 421b3–c6).

The ground has now been prepared for question (2) concerning the well-functioning of a polis. In Plato’s view Kallipolis functions well because it is organized on the basis of a principle of efficiency and quality, dubbed (by Nicholas White) the principle of “the natural division of labor”: “more and finer things are produced more easily when each man does one thing for which he is suited by nature, at the right time, being free from other pursuits” (*Rep.* II 370c3–5). This principle matches careers and vocations with natural talent and ability appropriately trained or educated. Plato thinks there is a natural hierarchy of such talent and ability symbolized in the myth of the metals by gold, silver, iron and bronze (*Rep.* III 415a1–7). When applied to this natural hierarchy the principle produces the tripartite social and political structure of Kallipolis in which every citizen has a place from which he is not to stray. Golden souls rule; silver souls defend; iron and bronze souls work; and the adage that a cobbler should stick to his last is heeded:

But we prevented a shoemaker from trying to be a farmer, weaver, or builder at the same time, and bade him remain a shoemaker, in order that the work of shoemaking would be well done; and similarly we assigned to each one of the others one occupation, for which he was naturally fitted and at which, being free from other pursuits, he was to work all his life, not letting slip the right moments for doing the work well. (*Rep.* II 374b6–c2)

The principle of the natural division of labor, it should be noted, is the root of Plato’s rejection of the two defining aspects of (modern and Athenian) democracy: freedom and equality (*Rep.* VIII 557a4, b4, 562b9–c2, 563b8). The freedom treasured by the democrat, as Plato recognizes, is the freedom to live as one wishes (*Rep.* VIII 557b4–10). But such freedom would be empty if human nature were as rigid as the Platonic principle implies, if each human being had the potential for just one vocation, just one way of life, rather than for a wide variety. Democratic freedom presupposes, contrary to the principle of the natural division of labor, that human nature is sufficiently plastic to allow for real choices among different lives. This assumption of human plasticity also lies behind the happy versatility in which Athenian citizens took pride (Thucydides, *Peloponnesian War* II 41.1), their ability to turn with ease from one occupation to another—from farming to bearing arms to ruling. The founders of Kallipolis, on the other hand, are so convinced that such doing of many things (*polupragmosunê*, *Rep.* IV 434b7, 9) precludes expertise—“Jack of all trades, master of none”—that they are prepared to back their principle of specialization by force, as the passage quoted above—“We prevented a shoemaker from trying to be a farmer”—makes plain (see in this regard Annas 1981, p. 79). As for the moral and political equality treasured by democracy, we see that the principle of the natural division of labor in conjunction with the myth of the metals denies both. The moral equality of human beings is at variance with the implication of the myth that some souls are worth more than others just as gold is worth more than silver, and silver more than bronze or iron. Political equality and the notion that goes along with it, that average citizens have sufficient intelligence to discuss and to decide public policy (*Prt.* 319b3–d7), are at variance

with the idea, entailed by the Platonic principle in conjunction with the myth, that ruling is an art requiring specialized knowledge attainable only by a few individuals naturally endowed with exceptional intellects. Socrates affirms both aspects of Platonic inequality at one stroke when he asserts that in Kallipolis “the better rules the worse (*to ameinon tou cheironos archei*)” (*Rep.* IV 431b6–7). Democracy, he complains, “distributes a sort of equality to both equals and unequals alike” (*Rep.* VIII 558c5–6). The total breakdown of the principle of the natural division of labor in a democracy produces in Plato’s view a kind of anarchy (*Rep.* VIII 562e3–5) superior only to the slavery of tyranny (*Rep.* VIII 564a6–8).

We come now to question (3) concerning the virtues of a polis. On the reconstruction of Plato’s argument that I have been offering the virtues of a polis, the qualities that allow it to function well, correspond to its various subfunctions. It is this matching of virtue to subfunction that explains, what is otherwise a mystery, why Plato assumes without argument that there are exactly the four virtues of wisdom, bravery, temperance, and justice. The function of a polis, it will be recalled, is to answer the needs of its citizens for wisdom, freedom, domestic tranquility, and wealth, happiness flowing naturally from the satisfaction of these needs. Plato thinks that Kallipolis functions well (is completely good) because he thinks these needs are fully and efficiently met by the cultivated and channeled natural endowments of its citizens. Its wisdom, residing in its rulers, provides wise policy; its bravery, residing in its warriors, preserves its freedom; its temperance—its “like-mindedness” (*homonoia*) or “concord” (*sumphônia*) of naturally worse and better as to which of the two ought to rule (*Rep.* IV 432a6–9)—prevents faction, or *stasis* (see *Rep.* IV 442c10–d1; and for *stasis* as the opposite of *homonoia* see I 352a7 and VIII 545d2); and its justice—each citizen doing his own—insures, among other things, that the workers by sticking to their jobs and not meddling in war or politics create the wealth a city needs.

“Doing one’s own,” as Socrates notes, is just another expression of the principle of the natural division of labor (*Rep.* IV 432b2–433b4). It should not be surprising that this principle, being the main condition of the well-functioning of Kallipolis, should turn up as one of its virtues. What is problematic is that this virtue should be identified with justice. Socrates is alive to this problem and offers four arguments in support of the identification. The first, the argument from residue, is that doing one’s own is the virtue that allows wisdom, bravery, and temperance to take root in a city, from which it follows that it must be a distinct virtue and hence identical with the only one left over, namely, justice (*Rep.* IV 433b7–c3). The second, or comparability argument, is that doing one’s own rivals wisdom, bravery, and temperance in its contribution to the virtue of a city, and no virtue aside from justice does that (*Rep.* IV 433c4–e2). The third, or juristic argument, claims that the jurors in a law court, in aiming at justice, aim “that neither litigant should have what is another’s or be deprived of his own” and links such having one’s own with doing one’s own (*Rep.* IV 433e3–434a2). The fourth is an argument from opposites: the meddling and exchange between the three classes, being the greatest evil that can befall a polis, is injustice; so doing one’s own, the opposite of such meddling and exchange, is identical with the opposite of injustice (*Rep.* IV 434a3–d1).

The prime question about these arguments is whether they establish that doing one's own is anything that is recognizable by us or by Plato's contemporaries as justice. The second and fourth arguments are of little help in this regard since they make no conceptual connection between doing one's own and justice. The juristic argument is better, appealing as it does to the use of "justice" in a legal system, thus connecting doing one's own to corrective or to penal justice. It has been objected that no ancient Greek juror would link having one's own with doing one's own and attempt to ensure the former by requiring the latter (Santas 2001, p. 91). But the juristic argument, like the other three arguments, applies not to historical Greek jurors but to jurors in Kallipolis (*Rep.* III 408c5–410a10); and, as Julia Annas reminds us, in Kallipolis having one's own and doing one's own do go together: "all have their own (that is, position, wealth, and honor are fairly and securely distributed) just because all do their own (that is, the basis of his society is one that reflects natural differences of endowment)" (Annas 1981, p. 120; see also Vlastos 1973, pp. 119–21; and Vlastos 1995, pp. 70–8).

As the juristic argument provides a reason from the standpoint of legal justice for identifying doing one's own with justice, the argument from residue provides a reason from the standpoint of distributive justice, a reason moreover that would be readily accepted by Plato's younger contemporary Aristotle. According to the argument from residue, doing one's own allows wisdom, bravery, and temperance to take root in a city. The way it does this is by assigning the tasks of ruling and bearing arms to those most qualified to perform them. From an Aristotelian perspective it distributes the most valuable of the apportionable goods, political office, among the citizens of Kallipolis according to the standard of wisdom. (Other standards are wealth and free status.) But if this is what doing one's own amounts to, then by Aristotle's theory of justice "doing one's own" is the expression of one conception (among others) of distributive justice—the aristocratic (*NE* V.3; *Pol* IV.8 1294a9–11; and Keyt 1991). One virtue of this interpretation of the argument is that it explains why Plato finds it natural to call Kallipolis an "aristocracy" (*Rep.* IV 445d6; VIII 544e7, 545c9, 547c6).

Whether the Platonic, or aristocratic, conception of distributive justice is a correct conception is a separate matter, dependent upon the correctness of the Platonic conception of a good polis; and this, of course, will be contested by both Athenian and modern democrats, who value freedom and equality and reject the principle of the natural division of labor upon which Kallipolis is based.

## 15.4 Psychic Justice

Plato now moves from polis to psyche and argues that the formula that defines justice in a polis also defines justice in a psyche: a psyche, like a polis, is just when each of its parts does its own. The passage justifying the transfer of the formula from one sphere to another has been variously interpreted, so it will be well to quote it in full:

Well, then, I said, [1] when one calls a larger and a smaller thing the same, are they unlike in that respect in which they are called the same, or like? Alike, he said. [2] And a just man will differ in no way from a just polis with respect to the form itself of justice, but will be like it. Like it, he said. [3] But a polis was deemed to be just when each of the three natural kinds (*genê*) within it did its own, and to be temperate, brave, and wise on account of certain other affections and states (*pathê te kai hexeis*) of these same kinds (*genôn*). True, he said. [4] And consequently, my friend, the one who has these same forms (*eidê*) in his psyche [5] we shall thus expect, on account of affections (*pathê*) the same as those, to rightly deserve the same names as the polis. Necessarily, he said. (*Rep.* IV 435a5–c3)

Focusing one's attention on (1) and (2) above, one might think that Plato subscribes to some principle of univocality, that he believes that a formula that defines a term in one application defines it in all applications. The problem with reading the passage this way is that Plato provides a counterexample to such a principle a few pages later in *Republic* IV itself. By his theory the formula that defines "just" when the term is applied to poleis and psyches does not define "just" when the term is applied to actions. A just action is characterized as an action that *produces and preserves*, in the agent, a psyche in which each part does its own (*Rep.* IV 443e5–6, 444c10–d1). By this characterization the formula defining a just psyche is a proper part of, and hence distinct from, the formula defining a just action.

The remainder of the passage above and the subsequent argument in *Republic* IV suggest a more subtle principle. Plato does not argue directly from the use of the same term to the applicability of the same formula; only after he has shown that polis and psyche have the same kind and number of parts does he define a just psyche as one in which each part does its own. His procedure indicates that he is assuming, not a principle of univocality, but a principle of similarity, or more precisely a principle of isomorphism: if a formula defines a term in one application, it defines it in all *relevantly similar* applications, similarity being understood as sameness of structure.

We can extract the general principle upon which his argument turns from (3), (4), and (5) above once we understand what Socrates means when he says that a psyche contains the same natural kinds (*genê*) or forms (*eidê*) as a polis. Kinds or forms are presumably different from parts (*merê*) since polis and psyche do not share their parts. (For the language of parts see *Rep.* IV 428e7; 429b2, 8; 442b11, c5.) What Socrates must mean when he speaks of the same natural kinds being in both polis and psyche is that polis and psyche have the same kinds of parts. Using the language of parts and kinds, we have two three-part systems, and three different kinds. One part of each system belongs to each kind. Thus, each part of one system has a counterpart in the other, part and counterpart being the parts that share the same kind. The general principle upon which Plato relies can now be expressed as follows: if (1) two systems have the same number of parts, if (2) the parts of the one system can be paired one to one with the parts of the other on the basis of the kinds to which the parts belong, if (3) these kinds of parts are the seat of certain affections (*pathê*), and if (4) the one system has a quality in virtue of its parts having such an affection, then (5) the other system has the same quality if its parts have the same affection. An affection, or *pathos*, in the context of the argument

is apparently a property, attribute, or characteristic (for this use of the word see *Prm.* 158e6–159a7).

This principle of isomorphism is used everyday in epigraphy. Suppose that an epigrapher transcribes a Greek inscription letter by letter from a stone tablet onto a sheet of paper, and suppose that he transcribes the unbroken string of capitals of the original by an unbroken string of lower case Greek letters. Since each letter of the Greek alphabet can be written as either upper or lower case, the corresponding characters of the inscription and transcription belong to the same kind: both Α and α are alphas, both Β and β are betas, both Γ and γ are gammas, and so forth. Many things will be true of the transcription that are not true of the inscription: it is written on paper rather than inscribed on stone, it was written recently rather than long ago, it is written in lower rather than upper case letters, and so forth. However, any sequence of letters that forms a word in the transcription forms the same word in the inscription; so a translation into English of the transcription will also be a translation of the inscription. This is an important fact if, as we may suppose, the transcription is more readily available and easier to read than the original.

The isomorphism of polis and psyche is supposed to resemble that of inscription and transcription (see *Rep.* II 368d1–7). Since it is far from obvious that a psyche has any parts at all, let alone the same number and kinds of parts as a polis, Plato mounts a long and elaborate argument to show “that there are the same [natural] kinds (*genê*), equal in number, in the polis and in the psyche of each individual” (*Rep.* IV 435c4–441c7). These natural kinds shared by polis and psyche, sometimes called “forms and characteristics” (*eidê te kai êthê*) (*Rep.* IV 435e2; see also VIII 544d6–e2), are three kinds of love: love of learning (*to philomathes*), love of honor (*to philotimon*), and love of money (*to philochrêmaton*) (*Rep.* IV 435e1–436a3 together with VIII 553c1). Within the psyche reason naturally loves learning, spirit honor, and appetite money (*Rep.* IX 580c9–581b11). Within the polis those who naturally love learning, honor, or money are respectively incipient rulers, warriors, or workers (*Rep.* II 374d8–376c6). When they are organized into three classes and properly trained and educated, the polis they constitute is just in virtue of each doing his own. The application of the principle of isomorphism is now straightforward: (1) polis and psyche each have three parts (2) of the same three kinds, which (3) provide the basis for doing one’s own, and (4) a polis is just in virtue of each part doing its own; hence, (5) a psyche is just if each of its parts does its own.

The problem with this argument is not the principle of isomorphism upon which it rests but the problematic psychological theory that Plato must adopt if the principle is to be applied. The wisdom lovers, honor lovers, and money lovers, who compose the parts of a polis, are agents with cognitive powers. If a psyche must have parts of the same kinds, they too must be agents with cognitive powers. Thus, Plato’s argument seems to demand that he anthropomorphize the parts of a psyche, that he conceive of reason, spirit, and appetite as three little men, or homunculi. This anthropomorphism is explicit in Plato’s two great similes of the soul, the composite creature (*Rep.* IX 588b10–e2) and the charioteer driving two horses (*Phdr.* 246a3–b4, 253c7–255a1), in each of which the psyche is depicted as consisting of multiple centers of consciousness; and it is implicit in Plato’s definitions of justice and the

other virtues. A polis and a psyche are just when each part of the polis or of the psyche does (*prattei*) its own (*Rep.* IV 441d5–e2). But to do its own a thing must act (*prattei*) and not simply move (*kinei*), which is to say that it must be an agent and not simply a faculty; and to act an agent must have cognition. The anthropomorphism of the two great similes is not simply metaphor.

It is important to bear in mind that, for the principle of isomorphism to apply, Plato's definitions of justice and of the other virtues must carry over word for word from polis to psyche. Political and psychic justice are not for Plato two species, or kinds, of justice but two applications of the very same concept: they are related as tall and short man, not as warm-blooded and cold-blooded animal. Plato's definition of temperance makes the general point crystal clear: "we should rightly say that this like-mindedness is temperance, this concord between the naturally worse and the naturally better as to which of the two is to rule *both in a polis and in each individual*" (*Rep.* IV 432a6–9). This means that anything presupposed by a definition of a virtue when the definition is applied to a polis is also presupposed when the definition is applied to a psyche. Since like-mindedness (*homonoia*) and concord (*sumphônia*) imply the sharing of a belief among the parts of a polis, the parts of a psyche must also have this capacity. And, indeed, when Plato, in discussing the virtues in a psyche, returns to the concept of temperance, he makes this implication explicit: "Isn't he temperate," Socrates asks, "because of the friendship and concord of these same elements, when the one that rules and the two that are ruled *believe in common (homodoxôsi)* that the rational element ought to rule and do not engage in faction against it?" (*Rep.* IV 442c10–d1). But if the parts of the psyche share beliefs, they must have cognitive powers. Plato's definition of bravery has the same implication. The bravery of a polis is the ability (*dunamis*) residing in one part of a polis to preserve a correct belief (*orthê doxa*) about what is to be feared (*Rep.* IV 429b7–c2, 430b2–5); the bravery of an individual is the same ability residing in the spirited element of the psyche (*Rep.* IV 442b11–c3). Thus the spirited element of the psyche has beliefs. Admittedly Socrates' definition of bravery in the individual, unlike his definition of bravery in the polis, does not mention belief explicitly; but, as we have just noted, this is not significant since the principle of isomorphism demands that the definitions be identical.

Of the many problems facing the sort of psychology that Plato is forced into by his use of the principle of isomorphism the most notable (and ironic) for a philosopher who stresses the importance of political unity is the problem of the unity of consciousness. The psyche as Plato conceives it has no center of consciousness; it is an harmonious or disharmonious committee of three. It is in need of an element that synoptically cognizes the actions and cognitions of its three parts, brings them to a focus, and acts for the psyche as a whole. Reason cannot perform this role since in the Platonic psyche reason is not always in the ascendant.

Plato implicitly recognizes the need for such an element when he describes the inner turmoil created by unjust action. He suggests that an embodied soul resembles a composite creature (man, lion, and many-headed beast) wearing a costume shaped like a man, and then continues as follows:

Let us say to one who asserts that it profits this man to act unjustly, but does not benefit him to do just things, that he asserts nothing other than that it profits him [1] to make the multifarious beast strong by feasting it, and also the lion and the things connected with the lion, [2] to starve the [inner] man and to make him weak, so that he is dragged wherever either of the other two leads, and [3] not at all to accustom one [creature] to another or make them friends, but rather to allow them to bite and fight and devour one another. (*Rep.* IX 588e3–589a4)

The “man” referred to at the beginning of the sentence is the composite creature dressed in its costume (the image of a soul dwelling in a human body). What is noteworthy is that the agency of this costumed creature is not reducible to the agency of its three inner parts: feasting the lion and the many-headed beast and starving the inner man are not actions of the lion, the beast, or the inner man. Nor is this agency due to the creature’s costume, the symbol of the human body. On Platonic principles bodies are totally inert and thus incapable of initiating action. All motion, and *a fortiori* all action, originates, according to Plato, in a soul (*Phdr.* 245c5–246a2; *Laws* X 894b8–896b8). Plato’s description tacitly posits a zoo-keeper who tends the menagerie of man, lion, and many-headed beast. The analogue of the zoo-keeper must be a psychic element distinct from reason, spirit, and appetite. I suggest that this element is the synoptic cognizer, or center of consciousness, that Plato’s psychology seems to demand on theoretical grounds.

My conclusion, then, is that Plato’s argument for his definition of psychic justice succeeds only at the price of a disjointed psychology of homunculi.

## 15.5 Just Action

Socrates says that justice resembles the principle of the natural division of labor, “though not in regard to the external doing of one’s own, but in regard to what is inside, to what is truly oneself and one’s own” (*Rep.* IV 443c9–d1). This idea, that justice is an inner state rather than a mode of action, is Plato’s climactic and revolutionary idea about what it is for an individual to be just. It is climactic in being the conclusion of a long argument extending over three books of the *Republic*; it is revolutionary in overturning an idea that seemed commonsensical then and still seems so today about the conceptual, or definitional, priority of *just action* and *just man*. The commonsensical idea, tacitly assumed by Polemarchus in his conversation with Socrates in *Republic* I, is that *just act* is conceptually prior to *just man*. Polemarchus claims that justice is giving to each his due (*Rep.* I 331e3–4), and Socrates takes this claim to imply that the just man is the man who gives to each his due (*Rep.* I 335e1–4). As this interchange makes plain, Polemarchus is tacitly assuming that “just act” is defined first and that a “just man” is a doer of just acts. Socrates thinks the conceptual, or definitional, priority runs in the other direction; he defines a “just man” as a man whose reason, spirit, and appetite each do their own and then defines a “just act” as an act that produces or preserves, in the doer of the action, this inner state (*Rep.* IV 441d12–e2, 443e4–444a2, 444c10–d1). An unjust act, on his theory, is one that destroys this inner state.



Since Plato is defining words of ordinary language, his definitions cannot depart too far from ordinary usage and still be regarded as correct definitions. Thus, it is important for him to test his definitions against the commonplace, or ordinary (*ta phortika*, *Rep.* IV 442e1). He must show that a man who is just, as he defines “justice,” will act, for the most part, as a just man would ordinarily be expected to act. He needs to show, in particular, that a Platonically just man will not do things that are ordinarily regarded as unjust. Consequently, just as he previously attempted to connect each citizen’s doing his own with the ordinary notion of political justice, he now attempts to dispel doubts about the transference of this formula from polis to individual by claiming that an individual each element of whose psyche does its own will act as a just person would ordinarily be expected to act: he will not embezzle a deposit of gold or silver, rob a temple, steal, betray his friends or his polis, break an oath or other agreement, commit adultery, disrespect his parents, or neglect the gods (*Rep.* IV 442d10–443b3).

That a person with the inner state of justice will not do such things is, however, just a bald assertion on Socrates’ part (*Rep.* IV 443e2–444a2) with nothing, in the immediately surrounding text at least, to back it up; and it is far from clear how the actions that Socrates enumerates fit his definition of “unjust act,” how stealing, betraying friends, committing adultery, and so forth destroy the inner state of justice in the soul of the doer. It has sometimes been thought that Socrates leaves the connection between action and inner state unexplained because no explanation is available, that there is a gap in his argument that cannot be bridged (Sachs 1963). Why, it is asked, must my conduct toward others affect the inner state of my soul? What prevents a person in whose psyche reason rules and the other psychic elements keep to their proper place from being a thief or an adulterer? Can no thief or adulterer be psychically healthy? (For injustice as a psychic disease see *Rep.* IV 445a5–b4.)

A charitable interpreter must seek answers to these questions and try to fill the gap in Plato’s argument. The distance he must travel to do this will determine the plausibility of attributing the filling to Plato rather than to the free imagination of the interpreter. Fortunately, in the present case the interpreter need not go beyond Plato’s dialogues. Most of the answer can be found in the *Republic* itself.

We can begin with Plato’s idea that the source of most evildoing is *pleonexia*, the desire for more and more, especially more and more money and more and more power. The nature and scope of *pleonexia* can be gleaned from Thrasymachus’ encomium of the pleonectic tyrant (*Rep.* I 343e7–344c8), Glaucon’s story of Gyges’ unrestrained *pleonexia* (*Rep.* II 359b6–360b2), and Socrates’ castigation of the pleonectic and bovine life of the many (*Rep.* IX 586a1–b6). In Plato’s view the state of the psyche of someone in the grip of *pleonexia* is like the state of the composite creature described in the passage quoted at the end of the last section; by feasting the lion and the many-headed beast while starving the inner man the zoo-keeper creates insatiable desires in the creature’s subhuman parts. The analogue of a just psyche is a composite creature in which the inner man is the strongest part; he (the inner man) fosters the tame heads of the many-headed beast while curbing its wild heads, and enlists the lion as his ally (*Rep.* IX 589a6–b6). Similarly, in a just psyche

reason is the strongest part; with spirit as its ally it fosters the necessary appetites and curbs the unnecessary ones, thereby purging the soul of *pleonexia* and removing the usual motive for theft, adultery, and other such crimes. (For the distinction between necessary and unnecessary appetites see *Rep.* VIII 558d8–559d3.)

This psychological analysis is only the beginning of a solution to the problem; it does not fully bridge the gap between a just psyche and forbearance from the acts on Socrates' list. Consider adultery. A man can be an adulterer without being licentious (*akolastos*): he can have a temperate sexual appetite for the wrong woman. What prevents a Platonically just man from being a temperate adulterer? Furthermore, adultery involves harm to others, to those who are betrayed. Surely, this consideration should play some role in the Platonically just man's forbearance from adultery. Finally, adultery (*moicheia*) is a legal concept and one that is defined differently in different legal systems. Adultery in Kallipolis where wives are held in common (*Rep.* V 457c7–461e9) is different from adultery in ancient Athens where a wife had a single husband and was required to be sexually faithful to him (MacDowell 1978, pp. 88, 124–5). Psychic justice must be moored somehow to positive law, law as actually laid down in a particular polis, if a Platonically just Athenian is even to be able to recognize what counts as adultery.

The issue is complex because in Plato's view positive law is often unjust (*Laws* IV 715b2–6). Only ideal law, law that is correct (*orthos*) according to the standard of nature (*Laws* I 627d3–4), is completely just. In the *Laws* the Athenian Stranger appeals to such a standard in passing judgment on the legal systems of the ancient world. Correct law, he claims, differs from faulty law in two respects: it aims at the common good rather than simply the maintenance in power of the established constitution; and it aims at the inculcation of all the virtues, not just one (*Laws* IV 705d3–706a4, 714b3–715b6). The laws of Sparta and Crete fall short of the ideal in aiming at victory in war and the bravery upon which victory depends while ignoring the other virtues (*Laws* I 625c9–626c5, 631a3–8; II 666d11–667a7); democracy, oligarchy, and tyranny are deficient in neglecting the virtues entirely and focusing only on maintaining the power of their rulers (*Laws* VIII 832d10–c7). The aforementioned constitutions can be ranked according to the degree of correctness (*orthotês*) or faultiness (*hamartia*) of their laws; and indeed the constitutional decline depicted in *Republic* VIII reflects such increasing faultiness. Timocracy, identified with the Spartan constitution (*Rep.* VIII 545a2–3), comes first after the ideal constitution; oligarchy precedes democracy because the miserliness of its rulers enforces a deviant sort of temperance, whereas even this caricature of virtue is missing from democracy (*Rep.* VIII 554b3–e6, 560c5–561a5); and tyranny comes last because of the tyrant's disrespect for law (*Rep.* IX 574d1–575a7).

There are two points about Plato's conception of correct law that bear on the Platonically just man's observance of the law concerning adultery. The first is that correct law, in aiming at the inculcation of all the virtues, is in Plato's view a form of moral education (*Rep.* IX 590c2–591a3, especially 590e1–2; *Laws* IX 857e4–5). Thus, obedience to correct law both produces and preserves psychic justice. But any action that does this is in Plato's view just (*Rep.* IV 443e4–444a2, 444c10–d1).

The second point is that correct law is an expression of reason. The connection of law and reason is a major theme of the *Statesman* and the *Laws*. In the latter dialogue the Athenian Stranger bids us obey the immortal element within us “giving the distribution of reason (*nous*) the name of law” (*Laws* IV 714a1–2), and in the former the Eleatic Stranger claims that laws are better or worse imitations (*mimēmata*) of the truth (*Pol.* 300b1–301a5). Though the theme is not so prominent in the *Republic*, it is there none the less. Socrates speaks of the tyrant fleeing law and reason (*logos*) (*Rep.* IX.587c2), declares that reason and law counsel a person to resist the pain of loss (*Rep.* X 604a10–b1), warns that pleasure and pain will be kings instead of law and reason if the pleasure-giving Muse of lyric or epic poetry is admitted to Kallipolis (*Rep.* X 607a5–8), and claims that what is furthest from reason is furthest from law and order (*Rep.* IX 587a10–11).

The connection of reason with law is understandable. Humans are embodied souls. That is why the psyche has its two lower parts, spirit and appetite (*Ti.* 69c3–72d8). As embodied souls humans are not self-sufficient, and their natural needs drive them to cooperate and to form poleis (*Rep.* II 369b5–7). Thus, if reason is to exercise foresight on behalf of the whole soul (*Rep.* IV 441e4–5), it must deal with these natural needs—the soul’s carnal appetitive desires—within a social and political framework. Recognizing the role of law and the common (*to koinon*) in binding a polis together (*Laws* IX 874e7–875b1; see also *Grg.* 507e6–508a4), reason wishes the soul of which it is a part to live in a polis in which law is respected and where there is friendship (*philia*) and a sense of community (*koinônia*) rather than faction (*Laws* III 695d2–3, 697c9–d1). Thus, if the law is correct and promotes the common, the person in whom reason rules has a strong motive to uphold it; and since he has no pleonectic motive to violate the law, he has no motive to be (like Gyges) a free-rider and benefit from the observance of the law by others while secretly breaking it himself.

But Athenian law is faulty. Does a Platonically just man have a motive to obey *it*? Here it is important to distinguish among the individual laws, and notice that the criminal actions that Socrates claims a psychically just man will not do would be proscribed by any legal code (Santas 2001, p. 61) and hence by the ideal code. (For the law on adultery in Magnesia, the imaginary polis of the *Laws*, see *Laws* VI 784e1–7.) In refraining from adultery one is obeying correct law, whatever sort of constitution one lives under. The gap in Plato’s argument can be bridged. The Platonically just Athenian will not be a thief, traitor, or adulterer.

A problem remains: will the Platonically just person obey faulty laws, particularly when his obedience will cause someone else to be treated unjustly? For example, will a Platonically just person, acting in an official capacity, enforce an unjust law or enforce an unjust application of a just law? Consider Socrates’ jailer. Socrates’ sentence, we may agree, is unjust (*Cri.* 54b8–c1). Would a Platonically just jailer administer the hemlock? This problem of the just executioner is a serious one for Plato because he appears to subscribe to three principles that are potentially conflicting: (1) that some laws are unjust (*Laws* IV 715b2–6), (2) that law should be strictly obeyed (*Pol.* 297d10–e2, 300e11–301a3), and (3) that one should never do anything that is unjust (*Cri.* 49a4–e3). He deals with a related problem in the *Crito*,

whether a just person should attempt to evade an unjust verdict of a legally constituted law court. But that problem is easier, from a philosophic standpoint at least, in that Socrates can avoid doing anything unjust by accepting his sentence of death. The harder problem is what leads Socrates to say that a person of reason (*ho noun echôn*) will not participate in politics in any except the ideal city (*Rep.* IX 591c1, 592a5–b1; see also *Ap.* 31c4–32e1).

## Bibliography

- Annas, J. 1981. *An Introduction to Plato's Republic*. Oxford: Clarendon Press.
- Cohen, S.M., and D. Keyt. 1992. "Analysing Plato's arguments: Plato and Platonism." In *Methods of Interpreting Plato and His Dialogues*, edited by J.C. Klagge and N.D. Smith, 173–200. Supplementary volume of *Oxford Studies in Ancient Philosophy*. Oxford: Clarendon Press.
- Dahl, N.O. 1991. "Plato's Defense of Justice." *Philosophy and Phenomenological Research* 51:809–34.
- Keyt, D. 1991. "Aristotle's Theory of Distributive Justice." In *A Companion to Aristotle's Politics*, edited by D. Keyt and F.D. Miller, 238–78. Oxford: Blackwell.
- MacDowell, D. 1978. *The Law in Classical Athens*. Ithaca, NY: Cornell University Press.
- Morrow, G.R. 1948. "Plato and the Law of Nature." In *Essays in Political Theory*, edited by M.R. Konvitz and A.E. Murphy, 17–44. Ithaca, NY: Cornell University Press.
- Reeve, C.D.C. 1988. *Philosopher Kings: The Argument of Plato's Republic*. Princeton: Princeton University Press.
- Sachs, D. 1963. "A Fallacy in Plato's *Republic*." *Philosophical Review* 72:141–58.
- Santas, G. 1985. "Two theories of good in Plato's *Republic*." *Archiv für Geschichte der Philosophie* 67:223–45.
- Santas, G. 2001. *Goodness and Justice: Plato, Aristotle, and the Moderns*. Oxford: Blackwell.
- Vlastos, G. 1973. "Justice and Happiness in the *Republic*." In his *Platonic Studies*, 111–39. Princeton: Princeton University Press.
- Vlastos, G. 1995. "The Theory of Social Justice in the Polis in Plato's *Republic*." In *Studies in Greek Philosophy/Gregory Vlastos, Vol. II: Socrates, Plato, and Their Tradition*, edited by D.W. Graham, 69–103. Princeton: Princeton University Press.
- White, N.P. 1979. *A Companion to Plato's Republic*. Indianapolis: Hackett.

# Chapter 16

## Plato on the Ideal of Justice and Human Happiness: Return to the Cave (*Republic* 519e–521b)

Yuji Kurihara

### 16.1 Introduction

In *Goodness and Justice* Gerasimos Santas (2001) succeeded in elucidating Plato's functional theory of good by examining Book I of the *Republic*. This is one of the greatest contributions he made to our understanding of the *Republic*. If Santas (2006, p. 137) is correct in saying that “the theory of function and virtue of Book I is used to give an account of justice in the rest of the work” (see also Santas 2001, pp. 105–6, n.16; *pace* Burnyeat 2002; Kraut 2005, pp. 451–4), we can make use of it to interpret crucial passages of Books II–X, as he often attempted. Interestingly enough, however, Santas seems hesitant to discuss a famous passage (519e–521b) in which Plato claims that the philosophers, after their education outside, must return to the Cave to rule (see 2001, p. 156). I believe the above passage should be read in the light of the functional theory, as Plato focuses on justice and happiness in relation to the philosophers' functions.

In this paper, following Santas's direction, I aim to answer two questions in 519e–521b with which many scholars have long struggled. (1) In what sense of justice is it *just* for the philosophers to return to the Cave? (2) Which do they increase by ruling, their happiness or their unhappiness? In Section 16.2, I reconstruct Plato's functional theory of good in Book I, relying on Santas's terminology. In Section 16.3, I focus on the first part (519e–520e) in detail to answer (1). In Section 16.4, I deal with the second part (520e–521b) to show that the philosophers become happier by ruling.

### 16.2 Plato's Functional Theory of Good

In the argument at the end of Book I (352d–354a), Plato intends to prove that the just person is happier than the unjust person. He bases his proof on the two ways of

---

Y. Kurihara (✉)  
Tokyo Gakugei University, Koganei, Tokyo, Japan  
e-mail: ykuri@u-gakugei.ac.jp

defining “function” (*ergon*) (352d9–353b1; see Santas 2001, pp. 66–75, and 2006, pp. 132–45):

*Exclusive Function:* X is the exclusive function of Y if and only if only Y can perform X.

*Optimal Function:* X is the optimal function of Y if and only if Y can perform X better than anything else.

Sight or seeing (*opsis*), for example, is the exclusive function of the eyes since one cannot see with anything other than the eyes. On the other hand, pruning vine branches is the optimal function of a pruning knife since it can trim them better than any other instrument.

Plato then defines excellence (virtue, *aretê*) as that which enables something to function well (353b2–d2). Accordingly, it follows that whatever has a function functions well by its own excellence (virtue), and badly by its own defect (badness, vice, *kakia*) (353c5–7). If, for example, eyes are deprived of their own excellence (virtue), they will function badly and one will not see well (353b14–c8).

Applying these definitions of function and virtue to the human soul, Plato attempts to prove that the just man is happier and better off than the unjust man (353d3–354a9). For living is an exclusive function of the soul. Since the virtue of the soul is justice, the soul lives well by justice. Thus, the just man who has justice in the soul lives well, that is, lives happily; the unjust man who lives badly is unhappy.

As Socrates himself admits (354b–c), this argument is far from satisfactory to him, since he does not know yet what justice is. On the other hand, if Plato’s subsequent definition of justice is supplied, one can make use of this theory of function even to understand the content of happiness for Plato; for this argument indicates that there is an internal or formal relation between justice as virtue and happiness as well-functioning. Let us now move to 519e–521b, keeping in mind this theory of function in the case of the philosophers returning to the Cave.

### 16.3 Beyond Social and Psychic Justice

One of the biggest puzzles in 519e–521b can be seen in Glaucon’s assent: “We shall be imposing just commands on men who are just”<sup>1</sup> (*dikaia gar dê dikaios epitaxomen*, 520e1). It is Glaucon that sharply objected to Socrates’ claim that the philosophers ought to return to the Cave to rule.<sup>2</sup> For Glaucon was supposing that it is not just at all for them to do so (*adikêsomen autous*, 519d8). Why has he changed his mind suddenly? What does he mean by “justice” here? Answering these questions will lead us to understand Plato’s notion of justice in this passage.

As is well known, in Book IV Plato defines justice as “doing one’s own work” in two ways. First, the city is just if it consists of just citizens who do their “own work” (433a–434d).<sup>3</sup> Second, the soul is just if each part of it does its “own work”; and an individual who has psychic justice is just, doing his “own work” (441c–444a, especially 441d11–e1).<sup>4</sup> It is highly controversial how just citizens and just

individuals are related to each other (see, e.g., Williams 1973; Ferrari 2003); I do not intend to discuss the issue here, but I only want to point out that we can find these two characterizations in our passage.

On the one hand, it is obvious that in the preceding passage (514a–519c) Glaucon was thinking of the philosophers as being “just” in the sense that they concentrate on studying Forms, letting each part of their soul do its “own work.” So they are psychically “just” individuals who are independent of their social role in the city. This is why Glaucon is surprised to hear Socrates’ commands that they rule in the Cave, which he suspected would prohibit them from doing their “own work,” that is, philosophy (IX 580c–583b).

On the other hand, Socrates goes on to remind Glaucon of the fact that the philosophers have been reared by the city so as to engage in ruling eventually (see IV 420b–421c, V 462a–466c).<sup>5</sup> This is how they differ entirely from another type of philosopher who grows up spontaneously (*automatos*, 520b2), indebted to none for his growth (*autophues*, 520b3). The capability of ruling in the city is part of their character, constituting their “second nature.” So they will be socially “just” citizens in *this* sense, performing their “own work” that consists in ruling in the Cave.

Thus, it turns out that the philosophers in question have two sides to their character. They are psychically “just” if and only if they engage in philosophical activities, since they do their “own work” as philosophers, whereas they are socially “just” if and only if they rule in the Cave, as long as they are doing their “own work” as citizens. It is noteworthy that these two sides represent two different ways of life that cannot together exist at the same time. I shall call the former “private life” and the latter “public life” on the basis of the well-known Greek dichotomy of *idiai* (privately) and *dēmosiai* (publicly).<sup>6</sup> Since the concept of *idiai* is understood as the privation of *dēmosiai*, these two concepts are mutually exclusive. So our philosophers cannot live these two lives at the same time, which leads them into a difficult dilemma as to which life to choose here and now. How can they be “just” in deciding to return to the Cave without stopping being psychically just?

As I understand, Plato replies to this question by directing Glaucon’s attention to the concept of “the whole life” (*bios*, 520e4, 521b1, b9; *zôê*, 521a4) that consists of both “private” and “public” life. In fact, Glaucon’s assent at 520e1 comes directly from his understanding of the ideal system of ruling in rotation, when Socrates asks: “will they refuse to share in the labors of state each in his turn (*en merēi*) while permitted to dwell most of the time with one another in that purer world?” (520d8–9; also *en merēi hekastô(i)*, 520c1). Glaucon then comes to realize that there is no conflict between private and public life in such a way that our philosophers can live both psychically and socially just lives harmoniously.

M.F. Burnyeat (2001, p. 8) extrapolates a formula of justice to cover both social and psychic justice, according to which “justice is exemplified by any system of elements working harmoniously together for the good of the whole and of each” (cf. Santas 2001, p. 157). This might explain why our philosopher’s whole life can be called “just,” in case it participates in the Form of Justice. For, justice of the whole life makes it possible for two elements of it, that is, private and public life to work harmoniously together for the good of the whole and of each.



Thus, Glaucon can call the bearers of this life “just people,” insofar as they succeed in doing their “own work” as a whole. For, conversely, if someone devotes herself to spending her whole life only on philosophy, she cannot be called “just” in the above sense, since she lacks the other component, “public life” (*pace* Beatty 1976b, pp. 140–1; White 1986, pp. 239–43). This also applies to the reverse case, *mutatis mutandis* (*pace* Aronson 1972, pp. 394–6; Vlastos 1981, p. 122, n. 31; Mahoney 1992). It is justice of the whole life that harmonizes both private and public life so that the bearer of it can do his or her “own work” in a strict sense.

In having Glaucon accept that our philosophers have two sides to their character, Socrates says they “are more capable of sharing both ways of life” (*dunatous amphoterôn metechein*, 520b8–c1). Needless to say, these two capabilities are philosophizing and ruling (see Adam 1963, p. 103), which stand for two distinct functions that characterize our philosophers as such.<sup>7</sup> If this is the case, the third type of justice serves to unify the bearer’s life as a whole, because of which it becomes possible for a virtue (or virtues) of these functions to come into being, since justice makes other virtues grow up and preserves them as long as it is present, according to Plato’s description of social justice in Book IV (433b–c).

In conclusion, we have good reason to believe that Plato in this passage introduces the third type of justice that enables the bearer of the whole life to do his or her “own work,” making it possible for each function of private and public life to function well. This concept of justice is, as it were, beyond the scope of the isomorphic analogy between city and soul.

## 16.4 The Ideal of Human Happiness

According to Socrates’ immediate application of the functional theory of the good at the end of Book I, since one of the exclusive functions of the human soul is living and since justice, as the human virtue, enables the soul to perform its function well, the just person will live well and so be happy. In this argument happiness is found in functioning well.<sup>8</sup> Assuming this internal or formal relation exists also between the virtue of the whole life and happiness, let us now try to explicate what the virtue of the whole life is, in the hope of clarifying the content of happiness.<sup>9</sup>

The text I want to focus on (520e4–521a8) reads as follows:

(1) If you can discover a better way of life than office-holding for your future rulers, a well-governed city becomes a possibility. (2) For only in such a city will those rule who are really rich, not in gold, but in the wealth that makes happiness—a good and wise life. (3) But if, being beggars and starvelings from lack of goods of their own, they turn to affairs of state thinking that it is thence that they should grasp their own good, then it is impossible. (4) For when office and rule become the prizes of contention, such a civil and internecine strife destroys the office-seekers themselves and the city as well.

Using antithetical construction here,<sup>10</sup> in (1) and (3) Plato talks about the practicability of the ideal city and its conditions; in (2) and (4) he accounts for the reasons why, respectively. The main contrast lies between two different beliefs (*oiomenoi*, 521a6), whether political office and rule can be the good for those who rule or at

most necessary constituents of their happy life.<sup>11</sup> Starting from the first set, (1) and (2), let us consider this passage carefully.

(1) mentions a condition under which the ideal city can come into being; and (2) tells us the condition is that the philosopher must become happy by ruling in the Cave. So it seems very clear to me that Plato does not depict our philosophers as self-sacrificing in their ruling<sup>12</sup>; rather he thinks that by doing so, they can perform their function well. Thus, I agree with most recent scholars who hold that the philosophers become happier by ruling instead of staying outside the Cave.<sup>13</sup> The reason, however, might be different,<sup>14</sup> which I shall explain below.

Let us look at (2) again, where it is said that the philosophers become “rich” by living a good and wise life. It is noteworthy that Plato calls their happy life “wise” (*emphrôn*, 521a4), which might show what virtue Plato takes to belong to the philosopher-rulers of the ideal city. The Greek adjective “*emphrôn*” is used only once here in the *Republic*, but the adverb “*emphronôs*” is used twice (III 396d1–2, VII 517c4), in both instances modifying the verb “*prattein*.” The last instance, in which Plato states that anyone who is to act wisely in private or public must have caught sight of the Form of the Good, is more interesting. As the phrase “in private life or in public life” (*ê idia(i) ê dêmosia(i)*, 517c4) suggests, one can live one’s life as a whole wisely by grasping the Form of the Good. This applies to our philosophers who have already been educated so as to know the Form of the Good and attain the single aim and purpose (*skopos*) in life, that is, happiness,<sup>15</sup> to which all their actions, public and private (*idiai te kai dêmosiai*), must be directed (519c2–4).

As we saw in Section 16.3, our philosophers’ functions are both philosophizing in private and ruling in public. So, according to the functional theory of good, the virtue of our philosophers must be wisdom (*phronêsis*) that enables both philosophy and rule to function well (*phronimôtatoi*, 521b8).<sup>16</sup> Since the whole life consists of private and public life, our philosopher’s life as a whole must be happy, insofar as they perform their functions well by wisdom (*phronêsis*) along with justice.

Next, let us try to specify the content of the philosophers’ happiness, assuming there is an internal or formal relation between virtue and happiness, as stated above. For this I refer to Book IV, 443c–444a, where in contrast to social justice Plato explains what psychic justice consists in and to what state of the soul it gives rise. To focus on the second question, Plato points out that justice in the soul makes the possessor become (i) “dear to oneself” (*philon genomenon heautô(i)*, 443d5; see Shorey 1930, p. 414, n. d) and (ii) “a unit, one man instead of many” (*hena genomenon ek pollôn*, 443e1–2; see Shorey 1930, p. 415 n. f). Thus, in Plato’s view, happiness in the soul consists in (1) self-love and (2) unity of the self. It should be noted, however, that the “self” in this context must refer to the individual or the soul that is independent of social role in the city, since Socrates is speaking of psychic justice in contrast to social justice (443c4–7, especially *eidôlon ti*, c4).

Difficult as it is to apply this account of happiness directly to our philosopher’s case, it will be possible to infer a conception of happiness from it, modifying the content of “self” in accordance with the bearer of the whole life. Since in our passage Plato thinks about a harmonious composite of private and public life, the “self” of

our philosopher cannot be seen only from one side of those lives. What then does he mean by “self” here? What does he think about the good life as a whole?

Let us now go back to (3) and (4) that were quoted above, where Plato explicates the bad life of those who are eager to rule in a city. Lacking their “private good things” (*agathôn idiôn*, 521a5), the so-called “office-seekers” indulge themselves in “public ones” (*ta dêmosia*, a5), thinking “it is thence that they should grasp their own good (*tagathon*)” (a5–6). Accordingly, they mix up private and public goods, because of their wrong conception of happiness, for they regard office and rule in the public realm as the good (*tagathon*) for them. As a result, they cause civil wars in which they end up destroying themselves and other citizens. According to this description, we can know not only that the “office-seekers” despise themselves without thinking seriously what is good for them, but also that they have lost their identity wholly by confusing their public life with their own good. Due to their self-disgust and loss of self-identity, they come to lose their whole lives completely.

By contrast, it is plausible that due to wisdom our really “rich” philosophers keep reflecting upon their happiness, correctly taking office and rule to be its necessary constituents. Differently from the “office-seekers,” our philosophers never start any civil war that destroys themselves and other citizens; instead, they are willing to share their “wealth,” that is, their wisdom, with others, establishing friendly relations and unifying the city as a whole. This must be the perfect and best form of the whole life as composite of private and public life.<sup>17</sup> To employ the explanation of Book IV, we can conclude that our philosopher’s happiness consists in self-love and self-unity, where the “self” signifies not the soul of the individuals (let alone the rational part of it) but the bearer of the whole life including both private and public activities.<sup>18</sup>

## 16.5 Conclusion

In this chapter I argued that justice in our passage is neither a psychic nor a social one but justice of the whole life concerning both private and public realms. Without this third type of justice, our philosophers could not live just lives as a whole in both the private and the public realm harmoniously. Further, I suggested that their happiness consists in self-love and self-identity, in which case the “self” of a philosopher stands for the bearer of his whole life, harmoniously unified by the third type of justice. This accords with Plato’s functional theory of good, in that our philosophers perform their functions, that is, philosophy and rule, well with wisdom and justice.

Finally, I shall close this paper by appealing to Book VI (496a–497a), where Plato contrasts the philosopher-king with the philosopher who lives only in private (496b4–5, c7–8). The latter type of philosopher, living in an unjust city, never engages in public activities, thinking that he would come to an untimely end without benefiting himself or others, but he lives a quiet life, dedicating himself to philosophy as his “own work” (*ta hautou prattôn*, 496d6). In describing such a private philosopher, no doubt Plato has in mind the historical Socrates (see *to daimonion*

*sêmeion*, 496c3–4; compare *Ap.* 31b4, c4, 33a2, b7, 36c3). By contrast, Plato goes on to refer to the ideal happiness of the former type of philosopher, saying that in the ideal city only will he himself perform his function fully well and together with his own happiness preserve the common happiness (*meta tôn idiôn ta koina sôsei*, 497a5). Accordingly, we have here good textual evidence to show that the philosophers' return to the Cave is not self-sacrifice at all (cf. Ferrari 2003, pp. 102–3, 107–8), but contributes to the perfection of their function as attaining private and public well-being.

## Notes

1. Translations of Plato's *Republic* are mainly based on those of Shorey (1930, 1935).
2. "Do you mean to say that [if we do not allow the philosophers who have received the education to learn the Good] we must do them this wrong, and compel them to live an inferior life when the better is in their power?" (519d8–9).
3. "The proper functioning of the money-making class, the helpers and the guardians, each doing its own work in the state (*hekastou toutôn to hautou prattontos en polei*), . . . would be justice (*dikaiousunê*) and would render the city just (*tên polin dikaian parechoi*)" (434c7–10); see also ". . . the state was just by reason of each of the three classes found in it fulfilling its own function (*to heautou hekaston en autê(i) [polei] pratein*)" (441d7–9).
4. ". . . each of us also in whom the several parts within him do each their own work (*ta hautou hekaston tôn en auto(i) prattê(i)*)—he will be a just man (*dikaios*) and one who does his own work (*ta hautou prattôn*)" (441d11–e1; my translation).
5. "You have again forgotten, my friend, . . . that the law is not concerned with the special happiness of any class in the state, but is trying to produce this condition in the city as a whole, harmonizing and adapting the citizens to one another by persuasion and compulsion, and requiring them to impart to one another any benefit which they are severally able to bestow upon the community, and that it itself creates such men in the state, not that it may allow each to take what course pleases him, but with a view to using them for the binding together of the commonwealth" (519e1–520a4). For Plato's view of character formation, see Kurihara (2001, pp. 38–41).
6. See Cohen (1991, pp. 70–97); cf. Arendt (1958). For example, Plato's Socrates in the *Apology* often characterizes his activities as "private" (*idiai*) in contrast with activities in public (*dêmosiai*).
7. For the equation of capability with function, see Santas (2001, p. 69). The capability of philosophy is their exclusive function because it belongs uniquely to their nature (*tôn oikistôn tas te beltistas phuseis*, 519c8–9), while the capability of ruling is their optimal function because they can rule best, though other people can rule too.
8. Even if, as Santas (2001, p. 58) says, "Plato does not define happiness in the *Republic*," we can expect his notion of happiness to be shown in his discussions about justice.
9. Most scholars suppose that in our passage Plato does not prove that ruling the ideal city makes the philosophers happier; indeed Kraut (1999, p. 237) says "Nowhere in 519–21 or in the rest of the *Republic* does he [Socrates] return to Glaucon's challenge and spell out how the philosopher does profit from holding political office."
10. "*ei men . . . , esti . . . gar . . .*" (520e4–521a4); "*ei de . . . , ouk esti . . . gar . . .*" (521a4–8).
11. Plato, of course, takes it that ruling is necessary (*anagkaion*, 520e2; cf. 540b4, d7–e1) for the philosophers to be happy; cf. Ferrari (2003, p. 29). There is a subtle issue of his terminology of "necessity" or "compulsion" in this passage (520a8, 521b7; cf. 499b5, c7, 500d5, 539e4), which I cannot discuss here; cf. Brown (2000).

12. Pace Foster (1936); Adkins (1960); Cross & Woозley (1964, pp. 101–2); Bloom (1968, pp. 407–8); Aronson (1972); White (1979, 1986, 2002).
13. Kraut (1973, 1999); Beatty (1976a, b); Irwin (1977, 1995); Mahoney (1992); Vernezze (1992), etc.
14. Vernezze (1992, p. 345), for example, argues that “a life that involves ruling is necessary for the philosophers who wish to satisfy their *eros* for the Forms” in light of the discussion in the *Symposium*; see also Irwin (1977, 1995). In this paper I want to focus primarily on 519–21 and secondarily on other passages in the *Republic*, aside from other dialogues. For a helpful criticism of widespread interpretations, see Brown (2004, pp. 277–80).
15. I regard this “*skopos*” (519c2) as happiness (cf. Irwin 1995, p. 387, n. 21 and p. 389, n. 5) rather than as the Good (see Adam 1963, p. 101, and cf. Shorey 1935, p. 139).
16. In Book VI (500b–d) Plato depicts our philosophers’ life on two sides. First, they try to establish justice in the soul, studying Forms in privately (*idiai*) independently of social activities, so that they become as orderly and divine as it is possible for humans to be (see 500d1–2). Second, in public (*dēmosiai*) they aim to educate other citizens to possess “ordinary social virtues” (*tēs dēmotikēs aretēs*, 500d9) such as social justice or temperance. It should be noted that these are two different stages that must not be mixed up. So I disagree with Kraut (1999, pp. 244–9) who attaches much importance to 500b–d, arguing that the philosophers’ imitation of the Forms necessarily implies educating other citizens. I rather think that it is up to their free choice whether they attempt to educate their fellow citizens.
17. In the subsequent discussion (521b1–11), Plato takes our philosopher-rulers to be engaging in “true philosophy” (*alēthinēs philosophias*, 521b2; *hoi hōs alēthōs philosophoi dunastai*, 540d3–4) as including political activity, contrasting it with “pure and contemplative philosophy” outside the Cave; pace Kraut (1992, p. 337, n. 34), (1999, p. 235–6); Beatty (1976b, pp. 545, 568); White (1986, pp. 26–7); Mahoney (1992, p. 272); Brown (2000, pp. 2, 6); Cooper (2004, p. 263). By ruling justly, our true philosophers are to have their own rewards (*timas allas*, 521b9; cf. 519d6, 540d7; cf. 592a1–4) that are different from the present rulers’ rewards (516c–d, 540d4–5).
18. There is another case in which Plato is interested in the whole life including private and public activities. In Book IX (579c–580a), he states that the worst life is that of the tyrannical individual who cannot live as a private citizen (*idiôtēs*, 579c6), but must in fact become a tyrant in the city.

## Bibliography

- Adam, J. 1963. *The Republic of Plato* (Vol. 2, 2nd ed.). Cambridge: Cambridge University Press.
- Adkins, A.W.H. 1960. *Merit and Responsibility*. Oxford: Oxford University Press.
- Arendt, H. 1958. *The Human Condition*. Chicago: University of Chicago Press.
- Aronson, S.H. 1972. “The Happy Philosopher.” *Journal of the History of Philosophy* 10: 383–98.
- Beatty, J. 1976a. “Plato’s Happy Philosopher and Politics.” *The Review of Politics* 38:545–75.
- Beatty, J. 1976b. “Why Should Plato’s Philosopher Be Moral and, Hence, Rule?” *Personalist* 57:132–44.
- Bloom, A. 1968. *The Republic of Plato*. New York: Basic Books.
- Brickhouse, T.C. 1981. “The Paradox of the Philosophers’ Rule.” *Apeiron* 15: 1–9.
- Brown, E. 2000. “Justice and Compulsion for Plato’s Philosopher-Rulers.” *Ancient Philosophy* 20:1–17.
- Brown, E. 2004. “Minding the Gap in Plato’s *Republic*.” *Philosophical Studies* 117:275–302.
- Burnyeat, M.F. 2001. “Plato.” *Proceedings of the British Academy* 111:1–22.
- Burnyeat, M.F. 2002. “Virtue or a Mug’s Game.” *Times Literary Supplement* 5176:4–5.
- Cohen, D. 1991. *Law, Sexuality, and Society*. Cambridge/New York: Cambridge University.
- Cooper, J.M. 2004. *Knowledge, Nature, and the Good*. Princeton: Princeton University Press.
- Cross, R.C., and A.D. Woозley. 1964. *Plato’s Republic*. London/Basingstoke: Macmillan.

- Ferrari, G.R.F. 2003. *City and Soul in Plato's Republic*. Sankt Augustin: Academia Verlag.
- Foster, M.B. 1936. "Some Implications of a Passage in Plato's *Republic*." *Philosophy* 11:301–8.
- Irwin, T. 1977. *Plato's Moral Theory*. Oxford: Oxford University Press.
- Irwin, T. 1995. *Plato's Ethics*. Oxford: Oxford University Press.
- Kraut, R. 1973. "Egoism, Love, and Political Office in Plato." *Philosophical Review* 82:330–44.
- Kraut, R. 1992. "The Defense of Justice in Plato's *Republic*." In *The Cambridge Companion to Plato*, edited by R. Kraut, 311–37. Cambridge: Cambridge University Press.
- Kraut, R. 1999. "Return to the Cave: *Republic* 519–521." In *Plato 2*, edited by G. Fine, 235–54. Oxford: Oxford University Press.
- Kraut, R. 2005. Review of *Goodness and Justice*. *Ancient Philosophy* 25:446–63.
- Kurihara, Y. 2001. "The Choice of Unjust Lives in Plato's *Republic*." *Skepsis* 12:37–49.
- Mahoney, T.A. 1992. "Do Plato's Philosopher-rulers Sacrifice Self-interest to Justice?" *Phronesis* 37:265–82.
- Santas, G. 2001. *Goodness and Justice*. Malden/Oxford: Blackwell.
- Santas, G., ed. 2006. *The Blackwell Guide to Plato's Republic*. Malden/Oxford: Blackwell.
- Shorey, P., trans. 1930. *Plato: Republic I*. Cambridge, MA: Harvard University Press.
- Shorey, P., trans. 1935. *Plato: Republic II*. Cambridge, MA: Harvard University Press.
- Slings, S.R. 2003. *Platonis Respublica*. Oxford: Oxford University Press.
- Vernezze, P. 1992. "The Philosopher's Interest." *Ancient Philosophy* 12:331–49.
- Vlastos, G. 1981. *Platonic Studies* (2nd ed.). Princeton: Princeton University Press.
- White, N. 1979. *A Companion to Plato's Republic*. Indianapolis: Hackett.
- White, N. 1986. "The Rulers' Choice." *Archiv für Geschichte der Philosophie* 68:22–46.
- White, N. 2002. *Individual and Conflict in Greek Ethics*. Oxford/New York: Oxford University Press.
- Williams, B. 1973. "The Analogy of City and Soul in Plato's *Republic*." In *Exegesis and Argument*, edited by E.N. Lee, A.P.D. Mourelatos, and R.M. Rorty, 196–206, Assen: Van Gorcum.

# Chapter 17

## Surpassing in Dignity and Power: The Metaphysics of Goodness in Plato's *Republic*

Christopher Shields

*In the realm of what is known, the Form of the Good is last and is hardly seen; but once it has been seen, it is necessary to conclude that it is in every way the cause of all that is right and fine. (Republic 517b7–c1)*

### 17.1 Is the Form of the Good a Form?

Encouraged by Glaucon to advance his conception of the Form of the Good in one of the central scenes of the *Republic*, Socrates demurs: he lacks knowledge, he intimates, and contends that anything he might produce by proceeding on the basis of opinion, even true opinion, would likely prove shameful and ugly. Undeterred, Glaucon presses his request, allowing that he would be satisfied should Socrates deign to discuss goodness in the way he had already discussed justice, moderation and the other virtues. In response, Socrates declines more forcefully still: so much would satisfy him too, he says, but he fears that any effort along such lines would earn him only disgrace and render him a subject of ridicule (506b). He does not abandon all affiliated endeavors, however. After refusing Glaucon's call for a direct investigation into the good itself, he agrees to tell Glaucon about "an offspring of the good" (*ekgonos te tou agathou*; 506e3), which he advances for consideration as what is "most like it" (*homoiotatos ekeinō(i)*; 506e3–4). This is the sun.

Though developing over a scant 120 lines of the *Republic*, the Analogy of the Sun has understandably commanded the interest of scholars and philosophers since the time of its composition: criticized by many as hopelessly obscure, admired by equally many as the height of artful analogical reasoning, appropriated by Christianizing Neoplatonists and others who perceived godlike agency lurking in

---

C. Shields (✉)  
University of Oxford, Oxford, UK  
e-mail: christopher.shields@lmh.ox.ac.uk



the Form of the Good,<sup>1</sup> and investigated by neutrally minded scholars in the modern era who have rightly seen that it provides crucial data regarding Plato's moral metaphysics, his normative theory of virtue, his political theory, his philosophies of education and acculturation, his epistemology, and, most tantalizingly, his ontology of Forms.<sup>2</sup> For it is clear that the Form of the Good has pride of place amongst the Forms, that it is somehow prior to the others, both explanatorily and ontologically. Ultimately, Plato even concludes: "In the case of things known, one is to say, then, that not only is their being known present to them because of the Good, but that both their existence and their being is present to them because of that, though goodness is not being, but is still further beyond, surpassing being in dignity and power" (509b6–10).

What could it mean to suggest that the Good surpasses Being?<sup>3</sup> Against what standard of measure might such a claim even be given content? What, fundamentally, is being asserted in this passage regarding the ontological status of the Form of the Good in Plato's *Republic*, both on its own terms and in relation to other Forms? If we find ourselves without approaches to these sorts of questions, we are left in a position no better than Glaucon, from whom Socrates' arresting assessment elicits what Plato portrays as a grotesquely comic outburst: "By Apollo, so supernatural a superiority!" (*Apollon, ephê, daimonias huperbolês*; 509c1–2).<sup>4</sup>

Scholars have understandably hazarded suggestions about Plato's intended meaning. Some have come to think that what he intends simply cannot be as he is represented in our best texts, that indeed our text must be somehow emended, perhaps, as Schneider urged, by the addition of a single word, so that Plato is made to say that not *only* is goodness being, but that it surpasses other beings in dignity and power.<sup>5</sup> That the text cannot stand as we have it may seem in some measure plausible—if we take it to mean that the Form of the Good *has no being*, i.e., that it does not exist at all. For surely Plato has been at pains to promote not only its existence, but its supreme and ultimate importance in the education of the Guardians, as the Form whose knowledge is prized above all others: "The Form of the Good is the most important thing to be learnt, as you have heard many times, that in virtue of which, when just things and the others additionally avail themselves of it, they come to be useful and beneficial" (505a2–4). Hence, one might conclude, if he is now represented as denying its existence *tout court*, surely something is amiss in the text.

Others understand Plato more moderately, accepting the text as it stands, while restricting the scope of any denial he might be making about the Form of the Good. One proposal, widely endorsed if variously motivated, has Plato accepting the *existence* of the Form of the Good, but denying that it is a Form. So Krohn asserts with what one may assume is a calculated abruptness: "Die *idea tou agathou* ist keine Idee. . . ."<sup>6</sup> So much will sound bizarre from a not too distant remove; it seems to invite the immediate rejoinder that the Form of the Good is, well, *a Form*. So dismissive a rejoinder, however dialectically unassailable, merely serves to obscure the surprisingly forceful reasons available to those who, like Krohn, reach this striking conclusion.

In fact, Krohn's view resonates with a broad range of scholars, including Joseph (1948), who offers a fairly rich set of motivations. He maintains:

The Form of the Good then is not one among the other forms, to which being belongs and which are the objects of knowledge. From one point of view, reality is exhausted in them. That which is good, and the goodness of it, are the same; for nothing of what is good fails to contribute to that goodness which consists in its being just all that it is. From another point of view, its goodness is something beyond everything contained in our description of what is good: for we describe it by running over its constituent parts, the Forms which are the various objects of our knowledge; and its goodness is none of these. This, which I think we can understand in principle, though we cannot verify it in a complete apprehension of the real and of its goodness, is what Plato means when he says that this goodness is *epekeina tēs ousias presbeia(i) kai dunamei*.<sup>7</sup>

Various strands come together in this sort of view: in some sense, Plato's view eludes us, though we can come close. The Form of the Good, suggests Joseph, is not one Form among others; it is, rather, more like the structure exhibited by all the Forms taken as a totality: the Form of the Good is not *a* Form, because it is an organizing principle of all Forms, that feature responsible for their intelligibility taken corporately or individually.

I will call such views, which can be surprisingly well-motivated, without prejudice: the *Not a Form View of the Form of the Good*. Now I can state my negative thesis: the Not a Form View is false. My positive thesis, then, is the converse, that the Form of the Good is a Form alongside other Forms; that it has whatever ontology each and every other Form has; that there is no special paradox consequent upon this being so; and, finally, that Plato provides a perfectly good, rather local but highly significant reason for thinking that the Form of the Good is not *ousia*, but is beyond *ousia* in dignity and power, namely this: while necessarily co-extensive with Being, the Form of the Good is not therefore to be identified with it. Instead, the explanatory role it plays vis-à-vis other Forms marks it as importantly distinct from Being.

## 17.2 Characterizing the Good Before the Sun

When disputing about the Form of the Good and its characteristics, scholars may seem to rush into a zone where Plato fears to tread. After all, we opened with the observation that Plato himself refrained from describing the Form of the Good directly, suggesting in fact that any attempt in that direction would render him a laughing stock. For all that, he does seem prepared to characterize it richly, if analogically, and is moreover less direct about denying his knowledge than his initial reluctance to discuss the matter might incline us to believe. When pressed by Glaucon, Socrates asks, "Does it seem to you right to speak of things one does not know as if one were a knower?" (505c1–2), where earlier he had said of the Form of the Good, more qualifiedly, "we do not know it sufficiently" (505a5–6), in which instance the "sufficiently" (*hikanōs*) presumably covers the sort of sufficiency required for providing a direct and discursive account of its Being (*ousia*). This is

the sort expressly required of the dialectician who, having grasped an account of the Being (*ousia*) of each thing, is able to provide that account to both himself and to others, though only to the degree that he has knowledge (534b2–4; cf. 506e3–4). All these claims are compatible with Plato's characterizing the Form of the Good. First, his knowledge may be scaled, so that epistemic access is partial and incomplete. Second, he in fact never disavows all knowledge directly.

This seems fortunate, since we do find Socrates saying a great deal about the Form of the Good, about its importance as an object of study, its relation to other Forms, about the role it plays with regard to various particulars, and to its surpassing power and dignity. We may safely group Plato's characterizations into three sorts: (i) those pertaining to its role in education; (ii) those differentiating the Good from popular candidates for identification; and (iii) those concerning its status as an ultimate object of desire.

In the passages leading up to the Analogy of the Sun, we find the Form of the Good introduced first with a view towards its role in the education of the Guardians, stressing its importance in procuring benefits expected from virtuous conduct of various sorts:

- (1) It is the greatest object to be learned (the *megiston mathêma*; 504e4–5, 505a2).
- (2) It is by their drawing upon the Form of the Good that just and other virtuous things<sup>8</sup> become useful and beneficial (*chrêsima kai ôphelima*; 505a4–5).
- (3) Without knowledge of the Form of the Good, even if we were to know things maximally, there would be no benefit for us—just as if we possess<sup>9</sup> something without its good (505a6–b1).
- (4) Still, if the Guardian in charge of our constitution knows these things, then our constitution will be perfectly ordered (506a6–b2).

This last passage (4) seems especially noteworthy inasmuch as the antecedent may or may not be envisaged as impossible. If, necessarily, we know other Forms only when we know the Form of the Good, then our knowing them maximally in its absence will in fact be impossible. If that is so, then we are induced to imagine a modal counterfactual which might nonetheless be perfectly correct and, indeed, true.

Those in the second set of pre-sun passages distinguish the Form of the Good from some popular and less popular accounts of its nature. Most people suppose the good to be pleasure (*hêdonê*), and others, more refined sorts, understand it to be knowledge or practical wisdom (*phronêsis*). They are equally mistaken:

- (5) Those defining the Good as pleasure are shown to be mistaken by the bare fact of there being bad pleasures (505c6–9).
- (6) Those who define the Good as knowledge cannot specify what the knowledge in question is knowledge of; when pressed, they fall back into circularity by suggesting that it is knowledge of the Good (505b8–10).

The two criticisms seem to be connected. If the hedonists respond in a conciliatory vein that the Good is defined not as pleasure without qualification but in terms of those pleasures which are not bad, then they are left with the thesis that the Good is to be defined by certain pleasures, to wit: the good ones. The circularity here overlaps but is not exhausted by the epistemological circularity encountered by the more refined. We might justifiably ask the hedonists how they know which pleasures are the good ones; or we might simply point out their lack of progress in defining the Good in terms of good pleasures.

Finally, in the run up to the Sun, Plato reserves a special and central role in human motivation for the Form of the Good:

(7) Every soul pursues the Good and does all that it does for its sake (505d11–e1).

No one, suggests Plato, wants the ersatz good, or what appears to be good without being really good. On the contrary, everything we do, we do for the sake of the Good, the real Good, preferring knowledge in such affairs and in this realm readily disdain belief (*tên de doxan entautha êdê pas atimazei*; 505d8–9), presumably in view of unreliability of belief for procuring what is sought. This might sound somewhat pedestrian, were it not for the fact that Plato is here actively contrasting the Good with just and beautiful things (*dikaia kai kala*; 505d5) where, he contends, most people are pretty well content to acquire and accept what seems to be the case, even if it is not so. While Plato's general descriptive contention may invite scrutiny, the contrast implied is noteworthy for our understanding of his attitude towards the Good: even a lazy person, content to wallow in meretricious attractions while admitting that they only seem to be beautiful without actually being so, will, he supposes, disparage seeming goods and turn away from them in favor of the real thing. In so speaking, he may be excessively optimistic, or unduly discriminating, though he succeeds in marking an asymmetry he observes between the Good and fine or just things: even the uneducated grasp, if dimly, the centrality of the Good. They divine that it is something, though are in difficulty with respect to its nature, and have not acquired the abiding confidence with respect to it which they have developed in other cases.

Because Plato approaches his discussion of the Form of the Good by first highlighting its indispensability in the education of those Guardians suited to rule, it is unsurprising that the three kinds of characteristics he initially highlights concern its relation to human conduct: (i) the first, (1)–(4), on its role in education, highlight its indispensability in the education of the Guardians who rule and the benefits it bequeaths to those in possession of its knowledge; (ii) the second, (5)–(6), pertain to mistaken identifications and definitions, in both cases in virtue of misguided attempts to reduce it to other, more familiar properties, pleasure and knowledge; and (iii) the third, (7), introduces it as an ultimate and exceptionless feature regulating human conduct, where being and seeming coalesce as in no other realm. The Good is what human beings want, real goodness itself, not a comfortable simulacrum; and it follows that anyone suited to rule will have a clear-eyed view of what it is that we all, ultimately, seek. Still, it is noteworthy that the characteristics ascribed to the

Form of the Good up to the point of the Analogy of the Sun all in one respect or another pertain to its relation to features of the sensible world and its occupants. So much paves the way for a radical reorientation in the Analogy of the Sun.

### 17.3 The Analogy of the Sun

When Plato introduces the Form of the Good in the Analogy of the Sun, he concentrates exclusively, or almost exclusively,<sup>10</sup> on its relation to a realm of objects which are knowable but not visible (507b9–10). Not content to draw the analogy relying upon his readers to tease out its intended tenor, Plato highlights various features of special concern to him. The list is long, but it is possible for our purposes to focus on just these seven, the first of which sets the terms of the analogy while the remaining six report its intended consequences:

- (8) In the realm of reason, the Good stands to the objects of reason as the sun stands, in the visible realm, to the objects of sight (508b13–c2).
- (9) As the sun provides the light enabling objects of vision to be seen and the power to see to the faculty of vision, so the Form of the Good gives truth and intelligibility, enabling the objects of reason to be known, and to the faculty of reason the power to know (507d11–e3, 508b6–7, 508e1–3).
- (10) As the sun can be seen, so the Form of the Good is itself an object of knowledge (508e4).<sup>11</sup>
- (11) The Form of the Good is responsible (*aitia*) for knowledge and truth (508e3).
- (12) Still, knowledge and truth are not identical with the Form of the Good, it being still more beautiful than they are (508e2–3).
- (13) As the sun provides not only the ability to be seen but also the generation, growth and nourishment to what is visible, so the Form of the Good provides not only being known to the objects of knowledge, but also their existence and being (*to einai te kai tēn ousian*; 509b6–8).
- (14) Still, the Good is not being (*ousia*), but surpasses it in dignity (*presbeia(i)*) and power (*dunamei*; 509b5–8).

Setting aside the last and most esoteric (14), each of Plato's intended consequences is reasonably clear, if subject to scrutiny. (10) merely points out that the Form of the Good, as an object itself in the realm (*topô(i)*; 508c1) of things known, is itself an object of knowledge (cf. 509d, *Timaeus* 52b–d). Given that not everyone knows the Form of the Good, and given that Guardians are expected to learn it as their most exalted object (their *megiston methēma*; 504e4–5, 505a2), Plato's contention must be taken as modal rather than factive. Moreover, and for the same reason, although the text of this passage is obscure and likely foul,<sup>12</sup> his contention in this regard is fully to be expected. If the Form of the Good must be learnt, then it is a potential object of knowledge.

In different ways, (9), (11), and (13) maintain that the Form of the Good is *responsible for* the attributes of other Forms. (9) tells us *inter alia* that the Form of the Good provides truth and intelligibility to the Forms, where the analogue is to sunlight. If that analogy is taken narrowly, as it seemingly should be,<sup>13</sup> his suggestion will be that the Form of the Good provides an enabling condition by making the medium suitable for cognitive content. (11) states baldly a thesis to be presently investigated, that the Form of the Good is responsible for knowledge and truth (*aitian d'epistêmês ousan kai alêtheias*; 508e3–4), while (13) moves the responsibility out of the narrowly epistemic realm and into the existential. According to (13), the Form of the Good gives existence and being (*to einai te kai ousian*; 509b7–8).

What sort of responsibility does the Form of the Good bear to other Forms? While it is natural and in some ways appropriate at this juncture to say that the Form of the Good is *causally* responsible for the features of Forms mentioned, it is also in another way problematic and unhelpful. In the first instance, to the degree that this sort of claim is apt, it mainly postpones a contentful question, as to the kind of causality envisaged. If it is, in Aristotelian terms, *formal* causation, then we will want to know how that is to be explicated in this rarified context. (Is, e.g., being even a formal cause of two?) Further, there is an immediately contentious claim as to whether the Form of the Good is in some sense a teleological cause of the other Forms. It is a common enough claim that it is<sup>14</sup>; but others, perhaps in view of non-equivalent conceptions of final cause, deny that any such claim is even intelligible.<sup>15</sup>

More importantly, however, if we stipulate that the Form of the Good is only, in some suitably specified and restricted sense, *causally* responsible for the features of other Forms foregrounded by Plato, then we foreclose too quickly on the possibility that Plato envisages a non-causal grounding connection for the Form of the Good. I mean that we can easily specify a range of non-causal dependency relations exhibiting the kinds of priority appropriately regarded as in the neighborhood of Form dependency. We think, e.g., that premises *license* conclusions; that proofs *establish* theorems; that rationality *explains* grammaticality; that a well-coiffed hairstyle *makes for* a glamorous appearance; that the arrangement of the parts of an automobile in one way rather than another *yields* a lower center of gravity; or that Socrates' drinking hemlock *brings it about* that Xanthippe is a widow.<sup>16</sup> Most of these relations are asymmetric; some are atemporal and others simultaneous; and none is in any obvious way causal. Any dispute regarding this last observation will only encourage a welcome reflection on what is meant by causal connection, beyond *grounding* or *explaining*, and by extension, then, to what is meant, more narrowly, by treating the *aitia* at 508e3 as causal, or more broadly, ultimately, by the suggestion that the Form of the Good is a cause of the other Forms.

We can begin to become clear about the kind of responsibility Plato has in view if we focus on those attributes for which the Form of the Good is held to be responsible. In this connection, it is helpful to follow Aristotle in distinguishing two kinds of attributes Forms manifest.<sup>17</sup> In a plainly eristic context, in *Topics* 137b3–13 (cf. 113a24–32, 144a14–22, 154a18–20), Aristotle distinguishes between what we may

call (with Keyt 1969), *ideal* and *proper* attributes. Ideal attributes are those properties a Form has *qua* Form. Proper attributes are properties a Form has in virtue of being the particular Form it is. Thus, Ideal attributes are shared by all Forms, and would include properties in the range of: *being intelligible, being immutable, being abstract, being eternal, being universal*. So, Ideal attributes are a subset of the necessary attributes of Forms; they do not exhaust that class, since it will also comprise logical and categorial features shared in common with non-Forms.<sup>18</sup> Proper attributes are trickier. Unless we are prepared to foist on Plato an immediate incoherence, we should not think of proper attributes as the properties the Forms are, with the result that every Form will be literally self-exemplifying. (The Form of Humanity will be a human and so be capable of becoming bored even though it is immutable, and so on.) Rather, a proper attribute for any Form F might non-tendentiously be defined as: being such that instances of F are F.<sup>19</sup> So, the Form Redness, while not itself red, is such that its instances are red. Redness thus has the proper attribute, being red, though it is not literally self-exemplifying; Justice is neither red, nor has the proper attribute being red. Any self-exemplifying Form, such as Being Abstract, of course has the ideal attribute and the proper attribute *being abstract*; it is just that it is one of its instances, and so by the definition of proper attribute, it is abstract; but it is not abstract in virtue of that proper attribute; on the contrary, it is abstract in virtue of its ideal attribute. From what has been said, it follows that every Form has the Ideal attribute of having the proper attribute which it does; that is, every Form F, *qua* Form, has as an Ideal attribute having the proper attribute F.

Thus defined, the definition of ideal and proper attributes absolves Plato of some of the one- and two-level paradoxes hurled his way in the *Topics*. More to the point, it helps explicate the kind of responsibility Plato may have in view: the Form of the Good is responsible for the Ideal attributes of other Forms; but, less obviously, it is responsible for their individual proper attributes as well, though perhaps by a distinct sort of grounding relation.

Starting with the ideal attributes, it makes ready sense to think of Forms as having the features they have *qua* Forms in virtue of their being good: each Form is, *qua* Form, immutable. This clearly coheres with what Plato says at *Republic* 380d–381b,<sup>20</sup> that the less mutable a thing is, the better it is. After surveying a number of organisms and artifacts, he concludes: “Whatever is in a good state, then, whether by nature or craft, or both, admits least of being changed by another.” The examples produced by Plato suggest that he takes inviolability and persistence as good-making features, both traits that Forms, *qua* Forms, exhibit. In this sense, at least, their being Forms at all involves their being good.

Should we say further that they equally owe their being the *most real* objects of all to goodness? After all, Plato says directly, in (13) that Forms owe their existence and being (*to einai te kai tên ousian*; 509b6–8) to the Good. Santas suggests so,<sup>21</sup> only to subject Plato to an unhappy conflation between superlative reality and superlative goodness. This seems not only unhappy but unwarranted. Santas follows Vlastos (1973) in finding a degrees-of-reality hypothesis throughout Plato’s works, locating it as especially prominent in the middle period theory of Forms. Armed



with this independently paradoxical thesis, it is easy to read Plato as suggesting that the Form of the Good provides not only existence but, so to speak, highest-level existence, just as it confers the ideal feature of being the best of its kind. Even if we do understand Plato as articulating and accepting a degrees-of-reality hypothesis (about which I am doubtful), it would not follow that the Form of the Good is responsible for the ideal quality of *superlative being* (whatever that may be) in being responsible for being. Rather, in providing for the ideal attributes of Forms, including their existence, it would follow, by whatever principles of hierarchy which might be adduced to yield degrees of reality, that Forms are at the top. In any event, nothing in the analogy thus far requires conflation, or any degrees of reality at all.

Taking all that together, we can appreciate how the Form of the Good can and should be thought as underwriting the ideal attributes of Forms: their having such attributes makes them good; so, they participate in the Form of the Good. So, it is correct that the Form of the Good is a Formal cause of the other Forms, though now this is given the following content: the Form of the Good is immediately responsible for the ideal attributes of Forms, and their participating in it is, or is akin to, *proprium* dependence, where the explanatory asymmetries and fecundities justify our understanding priority in the face of mutual entailment and necessary coextension.

What about the proper attributes of Forms? As I have defined them, the proper attributes of Forms neither trade upon nor implicate Plato in objectionable principles of self-exemplification. Every Form is such that its instances have the property it is, a harmless sort of property for Forms to have and not something which needs to be grounded in anything other than their being the Forms they are. If we suppose that Forms are *paradeigmata*—in either or both of two senses, as exemplars or as patterns<sup>22</sup>—then we may expect from them the purity or completeness of being F in virtue of which they can serve as such: exemplars are paradigms by realizing patterns in an especially clear or uncluttered way. Patterns in particular are regulative, or normative for their particulars, and thus are models without being exemplars; so much, at any rate, seems to be the only way to think of the Living Being at *Timaeus* 30c–d.<sup>23</sup> (This also seems to be the sense at 540a, where we are told that philosophers who know the Good itself shall use it as a pattern for the right ordering of the state and the citizens themselves.) Here then is a way in which we expect the non-ideal attributes of individual Forms to be good: they are patterns, norms proscribing ideals that are often (but not always) realizable to varying degrees. If the Form of Justice is analyzed roughly as “being such that one’s proper parts are harmoniously integrated,” then this is normative for its instances: its being perfectly just is not its being an exemplar of Justice, but in its capturing precisely what is essential to all and only just persons, practices and institutions.

Taken together, these suggestions provide reasonably straightforward ways in which both the ideal and proper features of Forms can be understood as good; they therefore also provide a variety of unadorned ways in which Forms can be expected to participate in the Form of the Good as one Form alongside them, in the way that Unity, Simplicity, Abstractness, and Being may be Forms alongside them. If that is correct, then we should regard the Form of the Good as one Form among

many other Forms, and as performing a function which other Forms perform for both Forms and particulars: they are those entities whose presence makes entities exemplify the properties they are.

## 17.4 The Not a Form View of the Form of the Good

Why, then, should so many scholars be attracted to the Not a Form View? In addition to the obvious fact that Plato marks the Form of the Good as exceptional, there are three sorts of reasons, not at all mutually exclusive, which have inclined scholars to suppose that its role vis-à-vis other Forms cannot be at all as just described. First, there is a series of relatively local linguistic motivations. None of these is terribly compelling. Second, there are two more synthetic appraisals of its role which seem to have implications for its ontological status. These reasons will prove to underscore what is plainly correct, that the Form of the Good has a unique role in Plato's metaphysics of Forms; but they provide no reason for thinking, as its proponents contend, that the Form of the Good is itself not a Form but rather a teleological structure of other Forms. Finally, and most interestingly, whatever its narrow linguistic implications, there remains Plato's striking and elusive locution that "goodness is not being, but is still further beyond, surpassing being in dignity and power" (509b8–10). Surely this propels the Form of the Good into its own realm?

Let us first consider three related linguistic observations. First, some render the absolute clause *ouch ousias ontos tou agathou* at 509b8–9 as: the Good is not "an Essence."<sup>24</sup> One might dispute whether we should render *ousia* in this passage as "Essence" or "Being" (I have preferred "Being"). The crucial question for the present purposes, however, concerns whether it should be treated as an indefinite count-noun and so rendered with an indefinite article. If the Good is not *an* Essence, and every Form is an Essence, then it would follow that the Form of the Good is not a Form. Still worse, if it is not even a Being, then it will not exist at all. One might suppose that such a rendering would, if anything, prove too much: surely, the Form of the Good exists; it is some sort of Being or other, whether or not it is a Form. Still, it would be entirely possible for the proponents of this line of thought to suppose that "Being" is here a categorial term, so that it meant, roughly, "substance." Then the translation would not prove too much. It would, on the contrary, prove just enough, that the Form of the Good is not a substance, as the other Forms, according to Plato, are. This would have the effect very nearly of collapsing the two translations: Plato's point would then be that the Form of the Good is not in the same category of being as the other Forms: it would not be a substance, or it would not be an essence.

Those encrusted in a remorselessly categorial caste of mind might then wish to know what it might be if it not a substance or an essence; we may, however, set that concern aside at present. The proposal seems to be that the Form of the Good transcends categorization, that it has no home within the narrow alleys of crabbed

taxonomy. We need not, though, conclude anything so striking. Rather, the pertinent point at present is the modest one, that nothing in what Plato says here requires such a rendering. On the contrary, Plato may be saying something weirder still, that goodness is without being or, as I shall presently urge, something less weird altogether. In either approach, the indefinite article may simply be set aside.

A second linguistic motivation returns us to Krohn (1876/1976), whose striking assertion that the Form of the Good is not a Form has at least the merits of arresting forthrightness. The reason he actually goes on to provide, however, seems to mistake the current clause. The Form of the Good is not a Form, he says, "because it has no *ousia*, but is rather a power."<sup>25</sup> Plato does not, however, suggest that the Form of the Good *has* no *ousia*, but rather that it somehow *is* not *ousia*. This seems fortunate because, once again, depending on where we go with *ousia*, this would be tantamount to saying that the Form of the Good is but lacks an essence, is but has no substance, or is but has no being. Yet if it is a power, then, it has at least being. Be that as it may, it is still more striking that it should be a *power*, since in that case, unless it is understood to be some sort of baseless disposition, it will be natural to ask to what this power belongs. There do not appear to be many contenders.

Still, the thought that the Form of the Good is a power, and not a Form, at least suggests an answer to the difficult question of what its status might be if we deny that it is a Form. A second sort of suggestion is more plausible, and much better motivated, not by narrowly linguistic data, but by a reflection on how the Form of the Good might be thought to function as regulative for other Forms. In his Oxford lectures of 1925,<sup>26</sup> Joseph argued that the Form of the Good could not be identified with any of the Forms, but rather, being constituted by them, was rather something like a *structure* of the Forms. His argument for this conclusion is rich and suggestive, meriting serious consideration. Plato suggests that the Form of the Good is something more beautiful than knowledge and truth, which themselves are already regarded as highly beautiful things (508e4–6). Joseph is rightly impressed with a kind of aesthetic dimension in Plato's conception of the Form of the Good. On its basis, he draws out a comparison between beauty and the Good, treating both as *systemic qualities* by contrasting them with what we may call *dissective qualities*. Consider the sweetness of honey. Down to a very small level of analysis, every drop of honey is sweet. It makes no difference which part we select, sweetness is dissective for honey, equally distributed across its composite parts. By contrast, aesthetic qualities tend to be not only non-dissective but also organic or systemic. Helen may be beautiful, though we might not ascribe that property to her pancreas. Similarly, though Mahler's Second Symphony may be transcendently beautiful, some discordant bits in the third movement, taken in isolation, sound entirely grating. Further, beauty is not only non-dissective in these cases: it is systemic. In the case of Mahler's Second, it is plausible to assume that its overall effect and beauty are enhanced and heightened by the discordant bits; without their presence the resolutions which follow would hardly be so poignant. Importantly, for Joseph, the level of appraisal is the whole, or the system. Goodness, like beauty, is organic.

Thinking now of the goodness in the *Republic*, Joseph suggests:

The expressions *auto to agathon* and *hê idea tou agathou* mean the same thing. But what I am trying to bring out can perhaps be made clearer, if we regard *to agathon* as goodness, *agathotês*, and *tou agathou* as genitive of *to agathon* which is not goodness, but the subject displaying goodness. That subject is the *noêtos kosmos*, the system of ideas, the eternal plan, perhaps we might say, imperfectly realized in the sensible world. Its goodness gives the reason for the details of it, and they are good because it is shown in them.<sup>27</sup> (Transliterations are mine.)

Looked at this way, the Form of the Good will have to be a feature predicable not of some part of the realm of Forms, but of the realm as a whole, displaying its vast and orderly interrelations, involving the necessary intermingling of Forms which Plato will later in his career come to emphasize.

It is a tempting thought; but one which should be resisted. To begin, if we take the analogue advanced by Joseph seriously and we relate it to Plato's own presentation of the matter, we will find him treating beauty sometimes organically and sometimes as predicable of the parts of a beautiful whole. It need not be disiective in order to avoid being systemic. Accordingly, just as, we may allow, the realm of Forms is beautiful, so too, claims Plato, is the Form of the Good itself: it is indeed more beautiful than some others, because of its exalted position (508e4–6). But if individual Forms can be beautiful, and we are to think of the Beautiful as akin to the Good, goodness too can be predicated of the constituent parts as well as of the whole.

The significance of this thought leads us back to the ontology of the Form of the Good. If it were necessarily and exclusively organic, then we might well be right to think that it is not a Form, but rather a teleological structure of the Forms, or a Power of the system of Forms taken *in toto*. It is not, however; and so we should not be disposed to think of the Form of the Good as transcending the Forms by being something other than itself a Form. Rather, it will be, as are all other Forms, interwoven with other Forms. To be sure, the Form of the Good will be prior to the other Forms, but it will not therefore be something (or nothing) which they are not.

## 17.5 Surpassing in Power and Dignity

Yet Plato *says* that the Good “is not being, but is still further beyond, surpassing being in dignity and power” (509b8–10). What, then, is his meaning? We may say that those inclined to follow Joseph have something to say here: Goodness is not being because it is an emergent property of all the beings there are. The Form of the Good is the goodness of the full system of Forms taken as a whole. From this perspective, the Form of the Good bestows upon all other Forms the goodness they manifest only insofar as they contribute to the totality of Forms which they, taken individually, partially constitute. It is their goodness, and as such, in its turn bestows upon beings the goodness they derive only insofar as they contribute to the totality which is good.

Here, though, strict attention to the terms of Plato's analogy suggests something simpler, something more readily understood, and something more striking in its

directness than all that would suggest. In four places, in two sets of parallel passages in the *Analogy of the Sun*, Plato cautions against an impulse we might feel to identify qualities which, he insists, must be thought of as distinct. First, just as light and sight, though sunlike, are wrongly thought to be the sun, so knowledge and truth are goodlike, but neither of them is the Good (508e). Second, as the sun provides generation to entities in the visible realm, so the Form of the Good provides both Existence and Being (*to einai te kai tēn ousian*; 509b7–8) to entities in the intelligible realm, even though it is not Being.

Why the necessary caution? In both cases where we might be tempted to conflate, we find ourselves facing a distinction between *provider* and *provision*. We have, then, in each case, a fact about priority which Plato does not want obscured. In the case of the Form of the Good and Being, the temptation is especially threatening, because Being and Goodness, in the realm of the Forms, are not only coextensive, but necessarily so. That is, necessarily, every Form is a Being and every Form is Good; if we think, then, that wherever—of necessity—we find Goodness we find Being and wherever—of necessity—we find Being we find Goodness, then the urge to identify will be strong indeed. The temptation is not idle, as indeed, the vast majority of metaphysicians have succumbed to it when they have treated necessary coextension as sufficient for property identification. Plato here, as elsewhere, shows himself alert to a mistake naturally made by those insensitive to asymmetrical explanatory relations amongst the necessarily coextensive.

Taking that together, Goodness goes where Being goes, and vice versa, converting without exception in the *topos* of Forms. Still, the Form of the Good is not Being, but prior to it, thus surpassing it in dignity and power. While Being is, the Good provides.<sup>28</sup>

## Notes

1. Nettleship (1901, p. 232) is a useful corrective: "In the *Timaeus*, the supreme power in the universe is described in a personal way, in the *Republic* it is described in what we call an abstract way. . . the 'Form of the Good' in the *Republic* occupies the place in regard both to morals and to science which the conception of God would occupy in a modern philosophy of morals and nature, if that philosophy considered the conception of God as essential to its system." Cf. Joseph (1948, p. 22).
2. Santas (1980/1999, pp. 273–4) aptly sums up the interlocking sources of interest offered by the Form of the Good. He says of it: ". . . the theory of the Form of the Good in the *Republic* is truly the centerpiece of the canonical Platonism of the middle dialogues, the centerpiece of Plato's metaphysics, epistemology, ethics, and politics, and even his theory of love and art. The Form of the Good serves his metaphysics by bringing into relief the very ideality of the Forms, the eternal order and stability of the entities that must exist if this world is not to be a 'vast sea of dissimilarity.' It serves his epistemology by bringing into relief the knowability of the Forms, the attributes some objects must have if there is to be knowledge. The Form of the Good serves his ethics and politics, and his theory of love and art, by bringing into relief the superlative goodness of the forms, the features that must be imitated if the imitations are to have any value. In his theory of the Form of the Good, Plato was truly the first grand philosophical synthesizer." In this chapter, I concentrate on only the first of these many strands, focusing on its *metaphysical* centrality of the Theory of Forms. Where I have occasion to part company with Professor Santas on some fine-grained points of this interpretation, I do so only against

the backdrop of a great deal of agreement with his masterful paper, “The Form of the Good in Plato’s *Republic*.”

3. Ferber (2005) provides an excellent critical overview of some of the most important recent discussions of this question. Although I do not agree with his linguistic and philosophical arguments for the view that the Form of the Good transcends Being, I have nonetheless benefited from his clear and forthright investigation of Plato’s difficult contention regarding the relation between the Good and Being.
4. There is a useful caution for Plato’s Greekless readers in the range of translations into English of Glaucon outburst:
  - Grube/Reeve (1992): “By Apollo, what a daemonic superiority!”
  - Shorey (1937): “Heaven save us, hyperbole can no further go.”
  - Lee (1974): “It must be miraculously transcendent,” remarked Glaucon to the general amusement.
  - Waterfield (1993): “It’s way beyond human comprehension, all right,’ was Glaucon’s quite amusing comment.”
  - Cornford (1945): “Glaucon exclaimed with some amusement at my exalting Goodness in such extravagant terms.”
5. Schneider (1874, p. 16).
6. Krohn (1876/1976, p. 146). Strikingly, the reasons he provides actually seem to support the stronger contention that the Form of the Good does not exist at all: “Die *idea tou agathou* is keine Idee, denn sie had keine *ousia*, sondern es ist eine Macht, die mit unserer Denkweise nur als die Gottheit begriffen werden kann.”
7. Joseph (1948, p. 23–4).
8. I take it that *t’alla* at 505a2 refers, via 504d5, to 504a5–6.
9. The translation in the text accepts the MS.’s indicative (*kektêmetha*) over the OCT optative (*kektê(i)metha*). Plato’s illustration seems intended to draw upon a familiar fact. If the good of a car is supposed to consist in its ability to transport its owners around, yet it fails to do so, then there is no benefit for them in possessing it.
10. Ross (1953, pp. 40–1) reasonably finds significance in this fact, as does Santas (1980/1999, p. 252). There are several counterexamples to this claim if it is to be taken literally, but they are mainly inconsequential: Plato mentions a connection between the realm of *nooumena* and *horômena*, in the context of setting up his analogy: the Good *begot* the Sun as its analogue (508b12–13); and he contends that the Form of the Good gives the power to know to reason, our reason, a faculty belonging to creatures in the visible realm (507d11–e3, 508b6–7, 508e1–3).
11. Reading *gignôskomenên* with Adam for the OCT’s *gignôskomenês*. The text here is, however rendered, obscure and probably corrupt.
12. Adam (1902, Vol. II, pp. 83–4) provides a succinct overview of some of the relevant possibilities and a brief survey of the literature to that point.
13. Plato presents colors as existing in the dark but unseen (507e2), while characterizing light as a “third kind of thing” (*genos triton*; 507e1) whose absence is compatible with the existence of unseen colors.
14. So Joseph (1948). Similarly, Irwin (1977, p. 273): “The Good is the formal and final cause of the Forms’ being what they are.” More generally, Irwin suggests (1995, p. 273): “The good, then, may be understood not as something independent of the virtues and other specific goods, but as the appropriate combination and arrangement of them. This is why Plato believes that the good is not a ‘being’ in its own right, but beyond being. . . .”
15. So Santas (1980/1999, p. 255 n. 15): “. . . I find it difficult to see how the Good is the final cause of the Forms’ being what they are since I would have thought that final causes are invoked to explain actions, activities, and movements, whereas the Forms are ‘at rest’.”
16. The example is from Kim (1993, p. 32), whose paper rightly concludes: “. . . causation, though important and in many ways fundamental, is not the only such relation. . . there are other determinative relations that deserve recognition and careful scrutiny.”

17. In following Aristotle, we also follow Vlastos, Owen, Keyt, and Santas, who have all picked up on Aristotle's observation, though in non-equivalent ways and to different ends.
18. Thus, Keyt (1969) will be too broad in treating ideal attributes as those "whose absence from a thing entail that the thing is not a Platonic Idea." This states a necessary but not sufficient condition on ideality.
19. This seems to be all that Plato has in mind in the *Euthyphro* when he describes the Form of the Holy as "that by virtue of which all holy things are holy" (6d–e).
20. This is noted by Santas (1980/1999, p. 256).
21. Santas (1980, 1999, p. 257): "...here we do have a conflation of superlative reality and superlative goodness of kind."
22. I follow Prior (1983) in distinguishing two notions of *paradeigma*: (1) a fine and courageous citizen may be an exemplar (*paradeigma*) of virtue, someone who manifests the features for which he is paradigmatic (so *Laches* 187a7); and (2) as a pattern which is inherently general, or universal, a structure common to many things and but not something normally or necessarily manifesting the pattern it is (the laws are *paradeigmata* for good conduct at *Protagoras* 326c8).
23. See Prior (1983, pp. 38–9) and Vlastos (1973, p. 304).
24. So Fine (2003); my italics.
25. Krohn (1876/1976, p. 122).
26. These were edited and published posthumously by H.L.A. Hart as Joseph, H.W.B. (1948), *Knowledge and the Good in Plato's Republic* (Oxford: Oxford University Press).
27. Joseph (1948, p. 21).
28. An early draft of this was paper presented at conference in honor of Professor Gerasimos Santas, where he kindly offered many astute corrections and encouragements. I was first alerted to the richness of Plato's positive characterizations of the Form of the Good by Santas, "The Form of the Good in Plato's *Republic*." This chapter is now a classic in its field, the best and most sophisticated treatment of its topic, and a *sine qua non* for anyone interested in thinking seriously about the Form of the Good. It is with great gratitude and pleasure that I offer this discussion of the Form of the Good to Professor Santas on the occasion of this volume in his honor. Another early version of this chapter was presented at Corpus Christi, Oxford, at a Seminar on Logic and Mysticism in Plato convened by Dominic Bailey. I thank Dominic Bailey, Terence Irwin, Gail Fine, Ewen Bowie, and Lesley Brown for their probing comments. A later version was presented to the Philosophy Faculty at Trinity College Dublin, where it received especially incisive comments from Vasilis Politis. The penultimate version was given at the University of Notre Dame. I am most grateful to the members of the audience on that occasion for their informed and thoughtful reactions.

## Bibliography

- Baltes, M. 1997. "Is the Good in Plato's *Republic* Beyond Being?" In *Studies in the Platonic Tradition*, edited by M. Joyal, 2–23. Aldershot: Ashgate.
- Cornford, F.M., trans. 1945. *Plato: The Republic*. Oxford: Oxford University Press.
- Ferber, R. 2005. "Ist die Idee des Guten nicht transzendent oder ist sie es doch? Nochmals Platons EPEKEINA TIS OUSIAS." In *Platon über das Gute und die Gerechtigkeit*, edited by D. Barbarić, 149–174. Würzburg: Königshausen & Neuman.
- Fine, G. 2003. "Knowledge and Belief in *Republic* V–VII." In *Plato on Knowledge and Forms*, edited by G. Fine. Oxford: Clarendon Press.
- Grube, G.M.A., trans. 1992. *Plato: The Republic*. Revised by C.D.C. Reeve. Indianapolis: Hackett.
- Irwin, T. 1977. *Plato's Moral Theory*. Oxford: Oxford University Press.
- Irwin, T. 1995. *Plato's Ethics*. Oxford: Oxford University Press.
- Joseph, H.W.B. 1948. *Knowledge and the Good in Plato's Republic*. Oxford: Oxford University Press.



- Keyt, D. 1969. "Plato's Paradox that the Immutable is Unknowable." *Philosophical Quarterly* 19:1–14.
- Kim, J. 1974. "Non-causal Connection." *Noûs* 9:41–52. Reprinted in his *Supervenience and Mind*, 1993, 22–32. Cambridge: Cambridge University Press.
- Krohn, A. 1876/1976. *Der platonische Staat*. Halle. Reprint: New York: Arno.
- Lee, H.D.P., trans. 1974. *Plato: The Republic* (2nd ed.). Hammondsworth: Penguin.
- Nettleship, R.L. 1901. *Lectures on the Republic of Plato* (2nd ed.). London: Macmillan.
- Prior, W.J. 1983. "The Concept of Paradeigma in Plato's Philosophy." *Apeiron* 17: 33–42.
- Ross, W.D. 1953. *Plato's Theory of Ideas* (2nd ed.). Oxford: Oxford University Press.
- Santas, G. 1980/1999. "The Form of the Good in Plato's *Republic*." *Philosophical Inquiry* 1: 374–403. Reprinted in Fine, G. ed., *Plato 1: Metaphysics and Epistemology*, 247–74. Oxford: Oxford University Press.
- Schneider, O. 1874. "Versuch einer genetischen Entwicklung des Platonischen *agathon*." Dissertation, Ritter-Akademie zu Brandenburg.
- Shorey, P., trans. 1937. *Plato: The Republic*. Cambridge: Harvard University Press.
- Strycker, É. 1970. "L'idée du Bien dans la *Republique* de Platon." *L'antiquité classique* 39:450–67.
- Vlastos, G. 1973. "A Metaphysical Paradox." *Platonic Studies*. Princeton: Princeton University Press.
- Waterfield, R., trans. 1993. *Plato: The Republic*. Oxford: Oxford University Press.

# Chapter 18

## Comments on Plato's Causal Explanation

D.Z. Andriopoulos

### 18.1 Introduction

Since scholars have raised the question concerning Plato's methods, and since understanding his methods is a necessary condition for understanding or reconstructing his thought, the scholarly concern (earlier and recent) with methods used or discussed in his dialogues is not surprising. Further, if this is an exegetical necessity, then central concepts crucial to discussions about causal explanations (as well as widely accepted accounts of these very same concepts) need to be thoroughly reexamined in light of pertinent inquiries, especially those carried out in the last decades.

As is known, the core of the problem and the quest for causal explanations is contained in *Phd.* 95ff, the interpretation of which is of immense importance for exegetical accounts of many issues in Plato's philosophy—ontological, epistemological, ethical, political and aesthetic. Auxiliary to this are several relevant but scattered passages, found mainly in the *Rep.*, *H.Ma.*, *Ti.*, and *Phil.*, which reinforce, enrich and advance, or even revise *Phaedo's* basic causal concepts. However, due mainly to semantically ambiguous terms in the text, scholarly procedures at times arrive at puzzling and logically vulnerable outcomes; this might also be due to the kinds of general philosophical presuppositions of each scholar (or group of scholars), or even a tendency to assimilate old theories to modern ones.

My discussion, then, might be best considered as a revisiting of some well-known issues, a clarification and reconstruction of certain scholarly positions, and an attempt to elucidate, wherever possible, Plato's general ontological and epistemological schemata strictly relying on the pertinent textual materials.

---

D.Z. Andriopoulos (✉)  
Aristotle University of Thessaloniki, Thessaloniki, Greece

## 18.2 The Concept of *Aitia*

*General.* If Plato used or assumed a causal theory, and aimed to give an explanation or account (*logon didonai*) of everything, then it seems most appropriate that one first discusses the main constitutive concepts of Platonic causal explanations: *aitia*, *eidōs*, *agathon*, and *methexis*.

Let me start with *aitia*, and at the outset point out that, since every kind of explanation relies necessarily on the concept of *aitia*, it is of the highest priority that *aitia* itself, in turn, is given an explanation.<sup>1</sup> This, I hope, will be confirmed in the rest of my paper. Although the core of the topic is contained in *Phd.* 95a, I consider it proper as a preliminary introduction to the topic to present the following theses that are scattered throughout Plato's dialogues:

- (a) I was tremendously eager for the kind of wisdom which they call investigation of nature. . . to know the causes of everything, why each thing comes into being and why it perishes and why it exists. (*Phd.* 96a 8–11)
- (b) It is necessary that all which comes into being comes by virtue of some *cause*. (*Phil.* 26e14, 3–4)
- (c) That which produces is nothing else but the *cause*. (*H.Ma.* 296e7–8)
- (d) Everything created is created of necessity by some *cause*; because it is completely impossible for something without a *cause* to obtain its genesis; anything which came into being necessarily came into being by some *cause*. (*Ti.* 28a 5–7 and 28c3–4)
- (e) The *cause* is creating (*to aition poioun*). (*H.Ma.* 297a 4–5)
- (f) In my second voyage I discuss the quest for *cause*. (*Phd.* 99d 1–2)
- (g) I am going to attempt to show the nature of this *cause*. (*Phd.* 95e 9–11)
- (h) For the *cause* of generation and decay must be completely investigated. (*Phd.* 95e 9–11)

All (a)–(h) except (f) are general, introductory views expressing Plato's belief that explanations require giving the reason (*logon didonai*). In particular, passage (a) mainly expresses the rather psychological tendency toward "inquiry for knowledge of nature" (*epethumêsa tautês tês sophias*. . . *peri phuseôs historian*), based upon causes; the term *historia* obviously refers to what happened in the past, to the generative starting points of some events and phenomena that are already completed, and perhaps to some that are still in process, that is, not yet completed; and since they are natural, they are classified as belonging to the realm of becoming (*gignesthai*). Passage (b) states Plato's firm contention that *aitia* is by necessity responsible for the genesis of all (*panta*) things; the statement is universal affirmative, and aims to be applied to both the dynamic-sensible and static-intelligible worlds. I think that this should be taken to imply that there is a generative cause of everything, even of the Forms, despite the fact they do not come into being. For *Agathon* is the Form of the Forms, the cause (*aitia*) of causes (*aitiôn*). A narrow interpretation would certainly lead one to think there is an inconsistency here. But can we see the problem

from the perspective of the development of Plato's thought? At any rate, it is very difficult to ignore the claim that *Agathon* is the cause (*aitia*) of everything (*pantôn*).

In passage (e), the terms *poioun*, *poiêsis*, and *poiein* point also to creation and creativity, to an activity or process. In passage (d), terms such as *gignestai*, *genesis*, and *genesthai* semantically point to a dynamic dimension. The same is true, and perhaps more strongly, of such terms as *gignomena* and *gignomenon* (still in process), *genomena* (completed events), *genesthai*, and, above all, *gignesthai*. But at the same time and in the same passages—and this seems to me important—words such as *ex anangkês*, *anangkaion*, *anangkê*, and *adynaton chôris*, introduce a deterministic terminology. One wonders whether this is a preliminary, gradual introduction to a deterministic ontological and epistemological framework for both worlds of Becoming (*gignesthai*) and Being (*einai*). If it is, then how can these two groups of terms, pointing to two different concepts and functions, be compatible with causal patterns used or assumed by Plato? Did Plato mean two different causal patterns corresponding to two different but parallel ontological realms? If so, it might be appropriate for scholars mainly inquiring into Plato's texts and other secondary sources to endeavor to identify or reconstruct both *static* and *dynamic* causal patterns. To this end, even if only a tentative conclusion about these matters is drawn at the end of these reflections, something will have been gained. Our inquiry will constitute a beginning of an understanding of Plato's views on causes; others might carry on the inquiry further and provide us with more detailed and definitive answers.

In order to proceed with my discussion, I find it most helpful to use Plato's own ontological and epistemological schemata as well as his terminology. Although it might sound quite traditional, I will intentionally try to avoid what I would call the *fallacy of modernization*. Permit me then to diagram here the schema of the Divided Line, and proceed level by level.

Types of world	Types of object	Ontic and epistemic levels	Types of cognition
<b>Intelligible world</b> ( <i>Noêta</i> ) BEING ( <i>EINAI</i> )	<b>Ideas or Forms:</b> The Good ( <i>Agathon</i> )	<b>Level D<sub>2</sub></b>	<b>Intelligence or understanding</b> ( <i>Noêsis</i> )
	<b>Other Ideas or Forms:</b> Beauty, Truth, Justice, etc.	<b>Level D<sub>1</sub></b>	<b>Knowledge</b> ( <i>epistêmê</i> )
	<b>More Ideas or Forms:</b> Mathematical Objects	<b>Level C</b>	<b>Thinking</b> ( <i>Dianoia</i> )
<b>Sensible world</b> ( <i>Aisthêta</i> ) BECOMING ( <i>GIGNESTHAI</i> )	<b>Visible things:</b> Animals, plants, man-made things	<b>Level B</b>	<b>Opinion or belief</b> ( <i>Pistis</i> )
	<b>Images:</b> Shadows, reflections	<b>Level A</b>	<b>Imagining</b> ( <i>Eikasia</i> )

*Level A.* This level encompasses images, shadows, reflections on water and objects with dense, bright and smooth surfaces (see *Rep.* VI 509d–510a, 510e, 516a–b). The questions to be asked are: (1) Are there any causal relationships among the items in Level A? (2) Are there any causal relations between these items and the objects and events of Level B? (3) What are the causes of their existence? There seems to be no clear, direct evidence in support of an answer to question (1); however, some kinds of dynamic, albeit in some cases peculiar, causal connections can be detected in Platonic texts. I see such a connection in *Rep.* 515b: the echo is causally related by the prisoners with a passing shadow.<sup>2</sup> In the same passage, the sound uttered by a prisoner is rightly connected with the produced echo (though the sound can be classified as an entity of Level B). Despite one basic difference between echo and shadow (the first is an acoustic and the second an optic datum), they share important similarities—both are representations that are distinct from and inferior to things in the Cave; they are both removed from the things that they represent or adumbrate—i.e., sound and object. It seems that another causal connection of this kind, depending on succession, co-existence and regularity, might be identified in 516c–d, where honors, prizes and commendations are offered to those prisoners who can guess the order of the coming shadows, (projected on the Cave’s wall), which of them comes first and which second. Going a little further in this passage, I would like to add that the function, having a dynamic character, reminds us of what are called “laws of association of ideas.” By saying in *Phd.* 73d–74a that on seeing Simias one is reminded of Cebes, Plato makes use of “the law of contiguity.”<sup>3</sup>

In our case here, echo and shadow appear in the same space and are thus associated, (albeit erroneously) by the “bright” prisoner, who creates a connection between the two. Further, such a causal connection enables the prized prisoner to predict which shadow will be and when it will be passing in the future. In *Rep.* 516d “[the prisoner] is best able to remember their [shadows’] customary beginnings, sequences and co-existences, and so is most successful in guessing. . .;” this anticipates modern thought and, according to P. Shorey (1935, Vol. II, p. 127) “is precisely the Humean, Comtean, positivist, and pragmatic view of causation.”

The word *eiōthei* indicates *habitude*; that is, by frequent observation<sup>4</sup> a causal connection is established by habit (*hexis*).<sup>5</sup> The term *aitia* does not appear in this passage. However, it is reasonable to suppose from contextual patters that causality is at issue. Thus, inferences can be drawn concerning causal relations among items, conditions, and conditional factors responsible for their existence and behavior (always in terms of the knowing powers of the laymen prisoners).

*Level B.* Plato’s general statement about the content of this level refers to animals, plants, and man-made objects.<sup>6</sup> Related passages are scattered in several dialogues; but the richest and most important for our discussion are in the *Phd.* and the *Rep.* Let me start, and in a preliminary way, with Plato’s views in the *Phd.*<sup>7</sup> and in particular with the stage of “rejection” or devaluation, or, as I prefer to call it, the stage of “catharsis,” in which the ordinary naturalistic theories of causality employed by most of Plato’s predecessors are assessed and found deficient. (The textual materials

discussed here are relevant to both Level B and D and, accordingly, will figure in our discussion of both of these Levels.)

First, Socrates states in *Phd.* (96a8–10) that as a young man he was enthusiastically concerned with the investigation of nature (*peri phuseôs istorian*); at that time, he was bent on knowing the causes of each thing (*eidenai tas aitiias hekaston*) and why each thing comes into existence, why it decays, and why it exists (*dia ti gignetai hekaston kai dia ti apollutai kai dia ti esti*) (96a8–b1). Then he presents a series of empirical examples: (1) the two opposites of heat and cold which, according to some, produce a fermentation responsible for the nourishment (*suntrephetai*) of animals; (2) cognitive activity that consists in changes in the blood, or air, or fire—a totally physiological or materialistic doctrine of previous philosophers; (3) the brain (*engkephalos*, reminding us of Alcmeon Krotoniates<sup>8</sup>) responsible for sensations (*aisthêseis*) of hearing, vision, and smell, which in turn furnish data to memory and opinion and which, again, become (*gignesthai*) knowledge (*epistêmê*)—also, a physiological and empirical theory; (4) the heavenly and earthly phenomena (96b–c); (5) the dramatized example of his own sitting in prison and conversing that is not due simply to the composition of his body (out of bones and sinews, their anatomical construction and functions of contraction and relaxation), but to his moral will, choice and decision to stay there; (6) the position of the earth below the heavens “by putting a vortex about it,” or by the support of air, or even by using a mythical Atlas who is more powerful, immortal and all-embracing (98c7–99c5); (7) the qualities of color, shape, and other such things as sufficient ingredients to *cause* a beautiful thing (100c8–d2); (8) again, the opposites of heat, cold, and the corresponding elements of fire and snow (103c9–d8).

At this point it would be proper to mention that Plato *seems* to “reject”<sup>9</sup> Anaxagoras’ theory of causality—the theory which postulates Mind (*Nous*) as being responsible for the arrangement and causing of all things—though Socrates at first was pleased to read the Clazomenian’s book. In fact, Plato thought that if *Nous* arranges things, then it should do it in a way which is best for them; in each case one must seek to identify the *best* and the *excellent* (*skopein. . . to ariston kai to beltiston*), which implies also that one will necessarily know the inferior (*cheiron*). As stated earlier, although Socrates had high expectations of finding in Anaxagoras the teacher of the cause of things (*tês aitiias peri tôn ontôn*) and of the “necessity of it” (*anangkaion*), he expressed his disappointment because the Clazomenian did not offer a theory to explain “what is best for each” (*to hekastô(i) beltiston*) and what is the common good for all” (*koinon. . . agathon*). The proposed causes—air, ether, water and many other absurdities (*atopa*)—seem to be unacceptable. Plato does not seem to object to *Nous* itself, but rather to Anaxagoras’ misuse of it. To call all these “causes,” according to him, is absurd (*aitia men ta toiauta kalein lian atopon*). In particular, referring to the connections between and function of the bones and sinews of a man, on the one hand, and, on the other, his will and choice, with respect to his sitting in the prison, he maintains that “whoever talks in that way [i.e., treating bones and sinews as causes]” is unable to make a distinction and to see “that in reality a cause is one thing and the thing without which the cause could never be a *cause* is quite another thing. . . most people, when they give the name of cause to

the latter, are groping in the dark. . . giving it a name that does not belong to it. . . they do not look for the power which *causes* things. . . as it is best. . . nor do they think it has any divine force. . . they give no thought to the good [*agathon*], which must embrace and hold together all things.”<sup>10</sup>

Second, in *Rep.* 527d, while discussing education, Plato refers to astronomy and its practical uses. In particular, he emphasizes things of everyday life, things which belong to Level B, whose existence as objects or events certainly has a dynamic character, especially in their causal interrelations—internal and external causal relations to things located at both the level of *gignesthai* and that of *einai*. The passage refers to calendar divisions—months, seasons, years—and to their use in everyday activities—agriculture, navigation and military art.<sup>11</sup> In this context the term *aitia* is absent; no direct reference appears. However, some indirect inferences might be drawn. Since the results of the arts of agriculture, navigation, and warfare are caused by several factors; and since one of them—the temporal factor—determines timing, which in turn determines positive or negative effects, the temporal factor can be taken as a necessary condition for these activities. Further and by reduction, time with its divisions and subdivisions is determined by astronomical factors, e.g., the planetary movements. These relations apply among material objects, movements and activities falling within Level B.

In *Rep.* 508c, juxtaposing the two regions (*topoi*) of the intelligible and visible (*noêtos* and *horatos*), Plato refers to (a) the eye (*ophthalmos*), (b) its difficulty in focusing on objects that are visible, (c) the colors (*chroas*), and (d) the daylight as causal factors that can negatively affect seeing. And in 508d, referring to the caused cognitive effects, by such causative factors, Plato states that: “. . .when it [psyche] inclines to that region which is mingled with darkness, the world of becoming and passing away, it opines only and its edge is blunted, and it shifts its opinions hither and thither, and again seems as if it lacked reason.” Also in *Rep.* 507e, by requiring a third kind (*triton genos*) as a necessary causal factor—that is, daylight as medium—he indicates that without its presence no visual effect is produced, even if the object and the colors are present; in other words, when the necessary and sufficient conditions are not met and the colors are invisible.

Regarding the third causal factor, we also find another reinforcing passage, where the necessity of the presence of a third factor, different from other necessary factors, is explicitly considered. For example, in the case of hearing, in addition to the sense and voice, air (although Plato does not say so explicitly) is needed, for without it no communication is possible: “the one will not hear and the other will not be heard” (507c). While describing the function of the sense of hearing in *Ti.* 67b, Plato is also seeking the causes (*aitias*) and the effects (*pathêmata*) of its activity; he is giving a definition of hearing by combining a chain of factors, without which no effect can be produced; these factors are necessary and sufficient for causing the perceptual result.<sup>12</sup> It is clear, then, that Plato is committed to several things about perception: (a) it is an activity, (b) involving a dynamic causal process, and (c) exhibiting a causal pattern that may consist in a (causal) chain whose initial (causal) link is remote and needs to be traced (see my 1988, pp. 114–15, and 2001). Although Plato’s use of the simile of the Sun purports to emphasize the status and



the role of the Good, one can also discern an important point concerning the variety in function among the causal factors for vision: sun, light, sense organs. The sun (and its light) is not vision, but one of its conditions; neither is it the *only* cause (*aition*) of vision.<sup>13</sup>

*Level C.* In this level belong various geometrical figures, the three kinds of angles, “and other things akin to these in each branch of science,” as well as arithmetical properties such as odd and even. These “are obvious to everybody,” and students of geometry and arithmetic are “treating them as absolute assumptions” and

make use of the visible forms, and talk about them. . . thinking of those things of which they are likenesses, pursuing their inquiry for the sake of the square as such, and the diagonal as such, and not for the sake of the images they draw. The very things which they would draw, which have shadows and images of themselves in water. . . The very things they treat in their turn as only images.<sup>14</sup>

The passage in *Rep.* 511a–b is also crucial for identifying the nature of the content of this Level:

This then is the kind [*eidōs*] of thing. . . [it is] intelligible. . . with reservation first that the soul is compelled to employ assumptions. . . not proceeding to a first principle because of its inability to extricate itself from and rise above the assumptions, and second, that it uses as images or likenesses the very objects that are themselves copied and adumbrated by the class below them. . . I understand, said he, that you are speaking of what falls under geometry and the kindred arts.

The following are some problems in Plato's discussion: (1) the nature, status and function of the five subjects (propaedeutic to dialectics (531c9–e2)) of arithmetic, geometry, stereometry, astronomy and harmonics, and (2) the causal connections among things in Level C as well as those in B and D, provided that textual evidence can be identified showing that Plato has in mind such connections. The five subjects can perhaps be called a bridge between the *changing* world of the senses and the *static* world of Forms. The demand for a logical structure requires securing a continuity of the Levels, consistency and strict logical interrelations, internal to each Level and across all Levels. Meeting this demand for a strict logical structure with hierarchical stratifications and necessary causal connections is a prerequisite to both inductive (particular to general) and deductive (general to particular) procedures; it is a logical condition for acquiring knowledge either for purely theoretical purposes or for practical applications to everyday activities, educational and others.

Plato uses several examples that refer to his favorite numbers and geometrical properties and measurements. He argues that the difference in their heights by a head, when two men are brought near, is not the cause of the superiority (inferiority) in size of the taller (shorter). Neither the division of one unit by another nor the addition of one to another causes the two, because the two operations are opposite (*Phd.* 97a4). In *Phd.* 100e4–101a, he again elaborates on this argument: (1) one cannot use the same thing (the difference by a head) as the *cause* of both the greater and the smaller; and (2) it is monstrous to claim that “one is great by something that is small.” In 101bc there is also rejection of two as the cause of “the excess of ten over eight.” The claim of the supposed contradiction is here repeated: “. . . if

one is added to one or if one is divided, you would avoid saying that the addition or the division is the *cause* of two?" (101b9–c2). Plato is maintaining that while material opposites can come into being from material opposites, this is not the case with other kinds of opposites; none of them becomes its own opposite either in us (*en êmin*) or in nature (*en tê(i) phusei*).

*Level D<sub>1</sub>*. It seems to me that the best way of dealing with this Level is to follow Plato's own voyage. The Athenian philosopher, after the "first voyage" (or semantic "catharsis," if one may describe it so) enters into the constructive "second voyage," "in quest of the cause."<sup>15</sup> In a very dramatic way he makes manifest his feelings of despair in trying to arrive at *firm* explanations of the processes and happenings in the material world.

Plato's intention in the second voyage is not quite clear.<sup>16</sup> As a consequence, different interpretations have been presented by scholars. One scholar, for example, claims that the second voyage must be taken as a mode "of inquiry about causation... [or] as way of discovering causes" (Rose 1966, pp. 464–73); and further, "the second-best way is explanation by formal causes and not the hypothetical method as such by which such explanations might be achieved. This type of explanation allegedly is second best to explanation by final causes, in the search of which Socrates was disappointed by Anaxagoras" (Sayre 1969, p. 4; Robinson 1953, p. 143, supports a similar position). A different interpretation by Zeller sees Plato's Ideas as functioning like Aristotle's formal, efficient, and final causes (Zeller 1922, p. 687 n. 1). According to another interpretation, the second voyage intends to (1) "establish his [Plato's] theory of causation" according to the procedures described in (100a), and (2) illustrate the method itself by showing how the theory of Forms as "causes" alone, in contrast to "the other learned causes" (100c) people speak of, "agrees with the hypothesis that the Forms exist" (Sayre 1969, p. 14). Robinson's own view (1953, pp. 143–4) of the hypothetical method is that it was set up to discover the "causes" of things in general—that is, to discover for any given type of thing the necessary and sufficient conditions for its being the type of thing it is.

But let us turn back to the text. Plato contrasts the causes he prefers with those ordinarily used, i.e., the "*learned* causes." His *superior* causes are the Forms themselves, which function in each case to explain the natural events. In 100b6 he tries to show "the nature of that cause" and clearly refers to absolute beauty (*kalon auto kath' auto*), good (*agathon*), great (*mega*), and all other such things to which he assigns existence (*kai t' alla panta*), a hypothesis that would help him to prove the soul's immortality. Plato's use of the term *hypothemenos* in his discussion of Forms as causes (*Phd.* 100b4–10) suggests that the existence of absolutes is assumptive. The absolutes are assumed as starting principles or points (*logoi*), and then the investigation proceeds. The procedure is briefly presented in this passage (100a3–7): "I assume in each case some *principle* which I consider *strongest* and whatever seems to me to agree with this, whether relating to *cause* or anything else, I regard as true, and whatever disagrees with it, as untrue." The absolutes comprise the lower section of Level D. Its second level, the highest place, contains the *Agathon*, the absolute of absolutes, the supreme cause upon which the entire knowledge and

truth are dependent. Plato is explicit: "you must conceive it [*Agathon*] as being the cause (*aitia*) of knowledge and of truth" (101c2–6). This is reinforced by his analogy in *Phd.* 99d6–e5, where the Sun's eclipse indicates the difficulty of obtaining knowledge and truth and, at the same time, the dominant role of the *Agathon*.

The causal significance of the Forms, which as hypotheses must be verified, depends upon the concept of *participation* (*methexis*). Illustrations of this can be identified in his subsequent argumentation: the existence of two is secured by its *participation* in the Dyad, i.e., the abstract entity which causally determines all the particular instances of two manifested in the world of Becoming: "anything can come into existence. . . by *participating* in the proper *essence* of each thing in which it *participates*, and therefore you accept no other *cause* of the existence of two than *participation* in duality" (101c2–6). Thus, it seems that cause is identified with participation (*methexis*)—a move later rejected by Aristotle. A second illustration offered by Plato refers to the causal relationship between absolute Beauty and the sensible qualities of shape and color. The presence (*parousia*) or communion (*koinônia*) of absolute Beauty (*auto to kalon*) makes an object beautiful, and certainly not its color or shape (100d5–9). The same claim is also made in *H.Ma.* (487c–d): beautiful things become beautiful by absolute Beauty. Next, he argues that participation of particular things in the abstract entities allows them to share with abstract entities the same name, and also shows the causal connection particulars (*hekasta*) have to kinds (*eidê*) (102b).

Again, the absolutes function as hypotheses by means of which we give accounts of why things are what they are and why they behave as they behave. Each hypothesis of the existence of an absolute will be appraised on the basis of its consequences. Plato was conscious of possible criticisms of the existence and adequacy of Forms as causes, and provisionally he suggests that

if anyone attacked the hypothesis [of cause as participation in the relevant Form] itself you would not pay attention, and refuse to reply to him, until you could examine the consequences of it, and see whether they are mutually consistent; and when you must offer an explicatory account of the principle you would pursue the same procedure: assume some other hypothesis (principle), the best and highest you could offer, and continue until you arrive at a more adequate one. You should not muddle things, like contentious men do, discussing at the same time the beginning and its consequences, if you wanted to discover some of the realities. (101d3–e4)

The entire discussion of Forms (*eidê*) as causes is part of Plato's effort to prove the immortality of the soul; the supposed causal efficacy of the Forms is used as a proof for the soul's immortality.<sup>17</sup> But within the context of this general discussion, Plato also purports to outline a theory of causality for explanatory uses. The use of the *eidê* as hypotheses is close to the demonstrative procedures in geometry, the discipline in which the term *hypothesis* originated. Perhaps it could be said that Plato was consciously imitating his contemporary mathematicians, as can also be seen in the *Meno*. A detailed discussion of Plato's concept of hypothesis, however, does not fall within the scope of the present study. Thus, I only wish to shed some light on what Plato means when he tentatively or provisionally puts forth the hypotheses that Forms exist and that they function as causes of objects and events.

*Level D<sub>2</sub>: The Agathon.* The concept of Platonic cause can also be examined with respect to carrying out a *diagnosis* of things that are present (*paronta*) and a *prognosis* of those which are to come (*esomena*). It seems to me that this is another aspect of the topic that deserves a detailed investigation. It suffices at this junction to say that knowledge of the real causes secures to a great extent the prognosis of the subsequent, expected events, those that are going to happen; and that this inevitably leads us to (a) the concept of *Agathon*, and (b) the implication that if the *Agathon* is the cause of the causes (*aitia tôn aitiôn*), then he who knows the *Agathon* can make both precise diagnoses and prognoses. Only he who “cognitively masters” the *Agathon* or sees the Sun is able to make right diagnoses and prognoses; and negatively, the prisoners in the Cave are unable to make right prognoses, if theirs are based upon the passing of regular time periods and projected on the surfaces of the Cave’s sides and not on the right diagnosis of *aitia*. Their life in the world of shadows limits their powers to reach the *Agathon*, the primary cause or cause of all the causes. On the other hand, the prisoner who escapes from the Cave, comes out and encounters the Sun, knows or knows somewhat the cause of causes, and despite his feeling blinded is able to make at least some prognoses that are free of error. The allegory of the Cave is particularly important (see Axelos 1956, pp. 95–8) in emphasizing Plato’s ontological and epistemological pre-suppositions, and the superiority of the *Agathon*, as the Form of Forms, and the only existential source of everything in both the worlds of Being and Becoming.

Four references should suffice to show the function of the Sun as the *aitia* of light and vision:

- (a) *Rep.* 508a–b: “[The Sun is]. . . the *author* and *aitia* of this, whose light makes our vision see best and visible things to be seen. . .” Again in 516c: “As the Good is in the *intelligible region* of reason and the objects of reason, so is this [Sun] in the *visible world* to vision and the objects of vision.”
- (b) 517b–d: “. . . this image [in the Cave]. . . we must apply, likening the region revealed through sight to the habitation of the prison, and the light of the fire in it to the power of the sun. In the region of the known the last thing to be seen and hardly seen is the idea of Good. . . [which] is indeed the *aitia* for all things of all that is right and beautiful, giving birth to the visible world, and the author of light and itself in the intelligible world being the *authentic source of truth* and *reason*, and that anyone who is to *act wisely* in private or public must have caught sight of this.”
- (c) 517d: “. . .if a man returning from divine contemplation to the petty miseries of man. . . appears most ridiculous, if, while blinking through the gloom, and before he has become sufficiently accustomed to the environing darkness [Cave] he is compelled in courtrooms or elsewhere to contend about the *shadows of justice* or the *images* that cast *shadows* and to wrangle in debate about the motions of these things in the *minds of those who have never seen justice itself*. . . do you think it at all strange?”

- (d) 518a: "But a sensible man . . . would remember that there are two distinct disturbances of the eyes arising from two *aitiai*: When the shift is from light to darkness or from darkness to light."<sup>18</sup>

In general, it can be argued that Plato's argumentation purports to (1) "reject" the prevailing causal explanations of reality, their objects, functions, and phenomena, thus preparing the ground for his own preferred (and not clearly stated) causal theory; and (2) elaborate, establish and use an enormous, lofty pattern based upon his ontological and cognitive multilevel world of what is, the worlds of Forms and sensible objects—the worlds of Being (*einai*) and Becoming (*gignesthai*). Since the theme is related to, or rather identified with, the world of Forms, and above all with the *Agathon*, and given the focal way that the *Agathon* is related with and determines the status of the Forms of the lower ontological levels, one cannot claim that Plato's concept of causality is clearly decipherable, easily identifiable, or even loosely definable.

At this juncture a special treatment of the *Agathon* is needed to show how it functions causally in Plato's entire grand ontological and epistemological schema, as presented in the simile of the divided line in *Rep.* 509–11, and furthermore of how it is used in his ethical and political systems as well in his theory of Beauty. A critique or even a plain review of the immense scholarship and variety of interpretations of Plato's views on the Good is presently impossible to discuss here for obvious reasons. On this occasion honoring my good friend Professor G. Santas, let me turn to his work and make a few comments on some of the basic points he puts forth in his important study (Santas 1980, pp. 374–403), "The Form of the Good in Plato's *Republic*."

After a highly scholarly and penetrating inquiry into the textual materials of Plato's Middle Dialogues, Santas concludes that the Form of the Good in the *Rep.* is truly "the centerpiece of the canonical Platonism in the middle dialogues, the centerpiece of Plato's metaphysics, epistemology, ethics and politics, and even his theory of love and art. . . [It] serves his metaphysics by bringing into relief the very ideality of the Forms, the eternal order and stability of the entities that must exist if this world is not to be a 'vast sea of dissimilarity' . . . In his theory of the Form of the Good Plato was truly the first grand *synthesizer*" (Santas 1980, p. 403).

Benefiting from some relevant works by other scholars, mainly those by G. Vlastos, D. Keyt and G.E.L. Owen, Santas proceeds to divide Plato's discussion into three rounds. In the first round he classifies and analyzes the relations of the Form of the Good with anything contained in the Intelligible and Sensible worlds, including all those connected with ethical and practical matters; the priority of the Form of the Good over all the Forms, as well as the objects and phenomena of the sensible world, is unquestionable. In the second round his discussion is, (1) limited to the relations of the Form of the Good with the Forms only, and (2) concerned with ontology and epistemology; the priority of the Good is taken for granted. Influenced by Vlastos' distinction between the dyadic scheme of *ideal* attributes (**IA**) and *proper* attributes (**PA**) of Forms, Santas explains how the Forms

are distinguished from the mathematical and other Forms in the lower levels of the hierarchical pyramid of ontology and epistemology. The Forms are identified on the basis of the (I1) or (I2) ideal attributes. Here the Good (*Agathon*) is taken as, (1) the *aitia* of the *knowability* of the Forms, (2) “of reason’s actually *knowing* the Forms,” and (3) “the *being* and *essence* (reality) of the Forms” (1980, p. 379). Further, Santas raises the question, “how can we understand the Form of *Agathon* as the *aitia* of ‘the being and essence of the Forms?’” His analysis, the structure of his arguments, and his examination of the textual evidence meet the highest standards of scholarship and are most convincing.

The third round focuses on the two upper levels of the Divided Line in the *Rep.*, where the *Agathon* occupies the top of the schema (509c–511e). Professor Santas draws a number of corollaries in his discussion which will be my main concern in the present discussion: the *Agathon* is the “cause” of the knowability of the Forms; their ideal attributes ((I1) and (I2)) make them knowable entities; and the *Agathon* is the *formal* cause of the *ideal attributes* of the other Forms: A knowable object, is “ungenerated, indestructible, not subject to increase and decrease, exists by itself ((I1) attributes), always is the same, the *same* in every respect, the *same* no matter compared to what, the *same* to all who apprehend it no matter from where ((I2) attributes)” (1980, p. 401).

More specifically, the third round has to do with the nature of the *Agathon*, its place in the hierarchical pyramid of the things that are (*onta*), and the way it determines the Forms, especially in the upper two sections (D<sub>1</sub> and D<sub>2</sub>). Eventually Santas focuses on the second round, raising crucial questions and pointing out the difference between the dialecticians and mathematicians with regard to hypothetical (mainly used by Euclid) and un-hypothetical reasoning; according to Santas, the first (dialecticians) deal with intelligibles (*noêta*), their attributes and relationships; the second (mathematicians), being “in an ambiguous epistemological position, deal with Forms and figures as images of Forms”; they are in a middle position and, using a neologism, perhaps one could call them “connectors” or “mediators.” The *Agathon*, thus, is the “cause” of reason’s knowledge of the Forms, whether mathematical or otherwise. In Plato’s ontological and epistemological schema, it is certainly the *unifying super entity*, making Plato “the first grand philosophical synthesizer” (1980, p. 403).

### 18.3 The Concept of *Methexis*

To understand what Plato meant when using the terms *aitia*, *aitios*, and *aition* and speaking of causal relations, especially in the upper realms of the Ideas, it is necessary to examine the concept of *methexis*. The term *metechein*—in its various forms as verb, noun or adjective—appears more than ninety times in Plato’s dialogues.<sup>19</sup> These uses vary among themselves, reflecting the diverse kinds of objects exhibiting the property signified by *metechein* or its cognates, as well as the different purposes for which the term and its cognates are used by Plato. For obvious reasons, I will



limit my attention to cases related to my attempt to interpret Plato's use of *aitia* and *aitiology*.

The inquiry in the textual material leads to the conclusion that *metechein* comes to be mostly used synonymously with "causal relating." *Metechein* has been translated in a variety of ways: "in virtue of," "participating in," "sharing in," "copying," "imitating," "responsible for," "resembling," "presence," "communion," "partake of," and so on. To my knowledge, most scholars have replaced "causally relating" with one or more of the above terms in their discussions of the topic.<sup>20</sup>

*Metechein* in various grammatical forms can be found in the following passages: *Phd.* 100e, 101b, b10, c; *Rep.* 455d, 432b, 510ab, 592a; *Ti.* 27c, 53c; *Chmd.* 158c4; *Prt.* 322c3–323d2; *Grg.* 485a3, 467e6; *Sph.* 255b, 256a, 259a. In all these passages, the role of *metechein* (in most of its forms) is causal, i.e., it generally replaces what is meant by: (1) various ontological connections among noetic entities (ideas), noetic entities and corporeal entities, corporeal entities and reflections or shadows (*eikasiai*)—that is, whatever is included in Plato's ontological schema; (2) epistemological connections and interrelations of concepts of the entities in (1); (3) linguistic connections among terms, propositions, and statements referring to the things in (1) and (2); and (4) in turn, connections of ontological, epistemological, and linguistic concepts with the *Agathon*.

But a question remains unanswered about the purpose of using the *metechein* in several such diversified contexts. Does Plato use it as a copula, as it is mainly used in ordinary language? One scholar contends that Plato purported to show three different meanings: (a) as copula (e.g., *metechei* means "shares in. . ."); (b) as identity sign (e.g., *metechei tou autou* means "shares in *sameness*"); and (c) as existential (e.g., *metechei tou ontos* means "shares in *being*"; Ackrill 1997, pp. 91–2). Understood in terms of causally relating, and taking into account the hierarchical stratifications in the various Levels, *metechein* is obviously incompatible with the identity sign; however, it is compatible with sharing in and sharing in being, but only in part, since things of the lower regions are determined by those of the upper regions and, eventually, by the *Agathon*: "[I]t is in virtue of participation in the Form of the Good [*Agathon*]. . . the other forms are 'always the same,' 'the same in all respects' . . ." (Santas 1980, pp. 374–403; reprinted in Fine 1999, pp. 267–74).

The passage is indicative of Plato's belief in an ascending *methexis* from the lower to the upper levels, but I remain puzzled as to how "the other Forms are always the *same* in all respects." Are the Forms the same among themselves? Or do they remain the same, unchangeable; and are they the same with respect to their causal connections with the Good? Are the ideal attributes (**IA**) the same in terms of the proper attributes (**PA**) of the Good? Can the (**IA**) of the Forms participate wholly in the Good's ideality? Or, is the participation of the Forms only partial and, thus, are their being and essences deficient? Santas' valiant attempt to answer the three questions, (1) "what is the essence of Forms?" (2) "what is the relation between the being and essence of the Forms and their knowability?" and (3) "how to understand the Form of the Good?", if successful, would not only clarify several crucial points, but also "make sense of Plato's position that the Good is the cause—in some appropriate Platonic sense of 'cause'—of the being and essence of the Forms."<sup>21</sup>



The phrase “appropriate. . . sense,” in my view, recalls my claim that *metechein* and “causally relating” are here semantically and functionally identical. The use of the two as synonymous—*metechein* and “causally relating”—applies in all levels and, thus, *Agathon* is a *Pan-aition*. To say that *methexis* does not apply between mathematical Forms and the *Agathon*,<sup>22</sup> one will have to deal with (1) the participation of two in the Form of the Dyad, and (2) the claim that the *Agathon* causally contains everything (*panta*). Why, then, there is no causal connection (*methexis*) between mathematical Forms and the *Agathon*? By saying that no pertinent textual evidence exists—that is, explicit references—we rather undermine the importance of several interpretations which we have considered here and are frequently defended. If textual evidence does not exist, one wonders, for example, how convincing the evidence corroborating the claim that the ideal attributes of Forms constitute (by *metechein*?) the proper attributes of the *Agathon* would be? Do not most scholars interpret on the basis of contextual implications or by presupposing properties, principles, and schemata? Why in cases in which there is absence of textual evidence are we ready to reject or doubt interpretations advanced by others but, at the same time, we employ similar ways ourselves?<sup>23</sup>

Commenting on this use of *metechein*, I wish to only raise questions which invite proponents of this kind of interpretation to answer: How by using *metechein* can one explain, for example, the relations among the *Agathon* and sensible objects and events? How, by using *causally relating*, can one explicate, for example, the connections among ideas and shadows or reflections? Do the denotative boundaries of the two concepts (*Agathon* and *causally connecting*) always coincide? Does *metechein* leave out cases of causal relations and *vice versa*? Do we have here cases of too broad or too narrow definitions?<sup>24</sup>

I wonder whether some interpreters, by reproducing what famous scholars in the past have written, take as a given that in Plato’s case, “cause,” “causing,” and “causally relating” can be easily interpreted as “participating in” or “sharing in.” Is it what Plato indeed meant by *metechein*, intending to fairly and convincingly explain causal relations only in the realm of the Ideas? Did he mean that *metechein* can cover all the causal relations and interrelations in all levels of his ontology?

Further, one wonders how *metechein* as causally relating is understood at the levels of both Becoming (*gignesthai*) and Being (*einai*), and how compatible is it with the commonly accepted notion that causation “is something which in some sense does something or other so as to produce or bring about an effect, which obviously refers to a concept of activity?”<sup>25</sup>

Some corollaries of the scholarly work by F.G. Hermann concerning the introduction and the philosophical use of *metechein* are worth mentioning.<sup>26</sup> His inquiry concludes that *metechein* (1) “as a term that denotes participation of a particular in a form is restricted to the *Phd.*, *Smp.* (211b1–5), *Rep.* (472c2–476d2), and the *Prm.*. . . but it is absent from the *Meno*, *Euphr.*, *Phdr.* and *Ti.*”; (2) “in the *Phd.*, Plato had used Anaxagorean language, with which the audience would have been familiar. . . and if one chose to employ that sort of terminology, one should be aware that the constituents of the world. . . are, in contrast with Anaxagoras’ analysis, not

corporeal"; (3) "Plato's *innovation* lies much less in the field of language and terminology than in the field of subject matter..."; and (4) "...Plato is... adopting and adapting their [the Pre-Socratic philosophers'] terminology while transforming their underlying analyses of the world."

## 18.4 Questions and Comments

(i) What are we to understand by "cause" and the Form of the cause in Plato (the *eidōs* of the *aitias*, *Phd.*100e)? The answer is extremely difficult, if not impossible. I have expressed my skepticism and concluded that no adequate definition can be drawn from the Platonic texts (Andriopoulos 1988a, b). I only note here that one scholar, despite his skepticism, says that although "writers on Plato typically decline to be very specific about this... perhaps wisely so," nonetheless goes on to suggest that "we can start from our uncontroversial paradigms," and asks whether Plato took "the Forms as causal on a par with these paradigms"; he then introduces for these paradigms the term "productive causes" and reshapes the question by asking if Plato "thinks of his Forms in... the *Phd.* and the *Rep.* as *productive causes*" (Bolton 1988, p. 92). But why is such a reluctance to be specific in defining the meaning of "cause"? If hitherto we have not reached some conclusions about what Plato understood by "cause"—Plato himself does not offer an explicit definition, and we cannot identify a tacit or presupposed one in his texts—then the observation that abandoning being specific about Plato's *aitia* is *wise* seems to me, on the contrary, to be *unwise*. It is rather wise to find out or to even infer in some contexts in the Platonic corpus what the Athenian philosopher meant by *aitia* and *aitios*. Attempts by G. Vlastos, M. Bunge, M. Frede and others offer a justification and ground for such a discussion about cause. Perhaps those relying on a traditional way of thinking will insist that, without a definition of *aitia*, no serious discussion about can be carried out; but such a strategy may not be the right one.

(ii) G. Vlastos and H. Cherniss, but especially the former, claim that the Forms are logical *aitiai*. According to Vlastos, Plato rather uses *metechein* ("sharing" or "participating in") of an object or event in a Form as a logical operation; a logical operation where the events and objects of the sensible world sufficiently satisfy the stated definition of the *ousia* of the Form. Does any beautiful object satisfy this type of requirement by having properties which to a sufficient degree correspond to the proper properties of the Form Beauty? If the definiens in the definition of the Form of Beauty is sufficient and constitutes its *ousia*, do we really identify for any beautiful *x* the *same* or *similar* properties? In what sense, then, does Plato take it that sensible instances satisfy the definition or account for the essence of the Form, if the sensible are deficient? Does the introduction of the term "metaphysical relation," used in a way that is *presupposed* by the fact that a sensible satisfies the definition of the Form, offer a convincing interpretation of Plato's scheme (underlying his accounts of cause and causal relations)? Certainly if the Forms are taken as logical *aitiai*, then they cannot be interpreted as efficient or productive *aitiai*

of the generation of their instances. However, no total exclusion of productive or efficient function of the Forms is meant by Plato; on the contrary, Vlastos himself points out that Forms functioning as productive *aitiai* are identified in the *Timaeus*, where the Demiourge acts to create the world; concerning efficient causes references can be made, not only to *Timaeus*, but also to *Phaedrus* and the *Laws*, where the concept of the self-moving soul appears (see Plato's discussion on self-moving or automatic psyche). According to Bolton (1988), Vlastos seems to conclude that we have to see here "an a priori physics according to which the laws of nature can be discovered a priori by analysis of the definitions of Forms," even if this physics is "a bad physics" (p. 93). Bolton goes on to say that, on the other hand, scholars such as J. Annas (1982, pp. 311–26), R. Hackforth (1955, p. 145) and I. Mueller (in J. Gentzler, pp. 67–91) contend that in the *Phaedo*, the Forms function as productive causes, and thus, Aristotle's criticisms are "appropriate."<sup>27</sup>

The disagreements are obvious. But one might think that these two contentions can be compatible if used in specific ways in each level, and not always expecting strict consistencies in Plato's thought. In the *Phaedo*, for example, causes may be considered as productive in the Anaxagorean sense; and as logical in the "second voyage" (*deuteros plous*), where the Forms and the *Agathon* are taken to be at rest. Again, I ask, given the plurality of perspectives in Plato's philosophy, why can't one accept a pluralism of contending interpretations?

(iii) Since for Plato the *Agathon* is the cause (*aitia*)—in fact, the cause of all (*aitia pantôn*)<sup>28</sup>: of Forms, mathematical, sensibles, events and processes that constitute the domains of coming to be (*genesis*), corruption (*phthora*), and being (*einai*), in both the dynamic (sensible) and the static (intelligible) worlds—seeing how Plato conceived its [the *Agathon*'s] nature is obviously both legitimate and necessary. Yet Plato avoids giving an answer to this question about the nature of the *Agathon*.<sup>29</sup> Plato's moving back and forth between the *Agathon* and the Sun, and at times the Cave, and relying on literary devices and style that exhibit exceptional poetic flair and unmatched imagination, cannot be considered a satisfactory answer to the question about the nature of the *Agathon*.

Here we have to take seriously into account Plato's confession that he cannot define the essence of the *Agathon*,<sup>30</sup> even if he tells us that in a realm exhibiting essences a definition is feasible (*Rep.* 534b8–c5). To me, this leaves us at an impasse, an *aporia* which, I think, should not be taken as a Socratic methodological step of pretending ignorance. As I said, to escape from this Plato invokes the Sun analogy and, in an auxiliary way, the Allegory of the Cave. These obviously suggest an absence of knowledge concerning the nature of the *Agathon*, which opens the way to "unwritten doctrines" and leads to apocryphism (Virvidakis 1989, p. 51). By "jumping"<sup>31</sup> to the Sun, replacing the *Agathon* by the Sun, "baptizing" it as the *aitia* of all kinds of things—of visibility of the visible objects, of vision in the dynamic world of Becoming (*gignesthai*), as well as of the knowledge of the intelligible objects (*nooumena*) in the static world of Being (*einai*), and finally of the being of all things (*pantôn*)—using the two (*Agathon/Sun*) synonymously, and rendering the *Agathon* as a kind of *pan-aitia*, is not a convincing solution—even if one takes

it in the mathematical sense as a super-axiom (Pappas 2006). Plus, as one scholar interestingly notes, the Analogy of the Sun (507a–509c) is responsible for “Plato’s univocity assumption in the case of Goodness. . . . Where [he] avails himself of both moral and non-moral senses of Goodness. . . .”<sup>32</sup>

The nature of the *Agathon* is, to some extent, defined in a *negative* way, that is, it is *not* being, or essence, or identical with truth and knowledge—an account that could be considered as being a “negative” definition.<sup>33</sup> On the other hand, given that, in a positive way, the ideal attributes of the other Forms are proper attributes of the *Agathon*, and that by *methexis* they are derivative from the proper attributes of the *Agathon*, the *Agathon* is causally responsible for their being. But to say that (or ask as to whether—see Santas 1988, pp. 45–47) the ideal attributes (being and essence) of the Forms are caused by the Form of the *Agathon* (the Grand Form of the Forms) we have to rely on the Platonic sense of the “cause,” and not ours.<sup>34</sup>

What then can be offered as a convincing exegesis of the Platonic position? Is it the case that Plato latently offers a *contextual* definition, a *functional* definition, or an *operational* definition of the *Agathon*? Perhaps one thinks that a definition is not needed, or Plato does not think it is needed, and/or he does not explicitly or implicitly use definitions. I am inclined to think that a definition is necessary, not only as the statement of the synthesis of the definiens (the essential characteristics of an object of knowledge), but also as a necessary tool for any cognitive endeavor. This is especially fitting in the present context; definitions for inquiries into the Platonic philosophical corpus, and especially for the topic under discussion, are necessary.<sup>35</sup>

To require definitions in order to understand and explain Plato’s conception of causal relations and causal interrelations throughout his entire ontological schema, it seems to me, we have to assume definitions for everything (noetic and sensible objects) in his schema; and all definitions should be of a static nature; in fact, we have to exclude the dynamic world of becoming (*gignesthai*) from definition; and further, we have to cognitively *discover*, not impose or create, a new ontology. But Plato accepts the dynamic world (even though depreciating it), and thus we cannot exclude it. Consequently, we cannot claim that, without definitions that capture the static essential properties of all objects (intelligible and sensible), causal relations (and interrelations) can be identified; for in the region of the sensible world (objects and events), we cannot secure changelessness and perennial causal relations; only relative definitions and temporary causal connections can be identified. Vlastos rather assumed a discovery of static definiens in the same way as it happens with the discovery of the so-called laws of nature. One wonders how invulnerable the “glorious” hitherto laws of nature are today.<sup>36</sup> The paradigm offered by Vlastos is not convincing, since the reliability of the laws of nature is revised or rejected decade by decade.

Further, I find it difficult to see how the essence of the *Agathon*, its being constituted by its proper attributes, is compatible with the *Agathon* itself *not* being essence but “transcending essence in dignity and power.”<sup>37</sup> Perhaps Plato had reasons for saying this, which might have included his belief that such a conception of the

relation of the *Agathon* to essence has many important ontological and epistemological implications. But still, this does not answer the question about the nature of the *Agathon* as the cause of everything (*aitia tôn pantôn*). The never-ending (*eisaiēi*) function of the *Agathon* as *pan-aitia*, in the Platonic sense of causality—even if taken as formal cause—leaves the question unanswered and the topic still problematic, despite the valiant attempts by G. Vlastos, G. Santas, and others.

One of my references is to G. Santas's (1980) introductory questions:

Why did Plato assign such a supreme position to the Form of the Good? What conception of goodness did he have, which allowed him to think of the Form of the Good not only as the final cause of everything that we do, but also as the cause of the knowability and even of the very being of his favorite entities, the Forms? And what connection did he see between the Form of the Good and mathematics? (pp. 374–5)

To these legitimate questions Plato has not given detailed and clear analyses or articulated comprehensive responses. And despite his demand for clearness and precision, he has rather left gaps that need to be discussed (see, in particular, *Rep.* 504d–e). It is not my purpose here to attempt to explain Plato's deficient or non-comprehensive analyses and responses to the questions raised by Santas.

(iv) Prognosis depends on diagnosis, since the former requires knowledge of causes or events and objects in both the visible and the intelligible worlds. The prisoners in the Cave, being conditioned by the continuous presence of the shadows in regular time periods, are able to prognosticate the appearance and order of the moving shadows on the wall behind their chained bodies. The lawgiver or the politician will mostly be right in his anticipation of coming events, if he first can make a successful diagnosis of the existing situation and can identify its causes; this, of course, presupposes regularity and repetition of events. Take, for example, Plato's circularity theory—the six types of governments, the right three, and the deviating ones. If the politician or lawgiver is to act properly, he must have mastered his causal theory—from the *Agathon* down to the first ontological Level, where shadows, images and reflections are causally determined by the sensible objects of the second Level, and in turn to those objects causally determined by their corresponding Forms. Only then will the lawgiver be successful in anticipating the coming things (*esomena*), and take the proper measures for achieving the good life for the citizens.

(v) I find it difficult to accept that Plato rejects the “learned causes” and even Anaxagoras' causal explanations. As Plato accepts the ontological Levels 1 and 2 in the pattern of the Divided Line, likewise he accepts, I think, this kind of causality as an explicatory instrument used by laymen and scientists of his time, just as it is used by today's scientists.

(vi) In general: (1) It seems to me that part of our puzzlement is often due to demanding, with unnecessary zeal, that Plato or Aristotle, who wrote some twenty-five centuries ago, should use our interpretive schemata, categories, or concepts. Perhaps a psychological tendency leads us to the urge to pseudo-modernize some ancient thinkers. Permit me to call it the “Fallacy of Modernization.” An example is

found in Gail Fine, who introduces terms such as “holist,” “coherentist,” and “foundationalist” to characterize Plato. (2) Why should we presuppose that in writing his dialogues Plato necessarily had in his mind definite concepts about Ideas and sensible objects, their functions, and their interrelations, as well as a definite place for them in an austere ontological and epistemological structure? The controversy, for example, as to whether the Forms should be taken either as logical, or productive causes, or whatever else, and the insistence that this was what Plato meant, seem to me to be such one-sided interpretations that they do not render philosophical justice to the Athenian philosopher. Why, then, do we silently presuppose either that Plato was highly clear about his concepts, or that he was always correct, logically consistent, and invulnerable in his thinking?

(vii) I claim that schemata, criteria, terms, synonyms, etc., drawn from the Platonic corpus are more effective, scholarly proper, and legitimate for semantic elucidations and evaluations. It is, I think, fallacious to introduce, after many centuries, philosophical neologisms, contemporary models, concepts, categories, and unjustified dilemmas and impose them upon the ancients' thought with an unnecessary zeal and anti-scholarly “demand.” This also applies to the hyperbolic use of the instruments and techniques of the apparatus of symbolic logic. I am pleased that Santas talks about caution, adopting Benson Mates' warning:

As Professor B. Mates has pointed out to me, one must be cautious in using the apparatus of symbolic logic to represent the structure of Socrates' arguments. . . the use of the horseshoe for “if. . . , then. . .” sometimes changes the strength of the assertion and quantification into oblique contexts, following “thinks that. . .” can give rise to paradoxes. I have tried to be cautious in these matters, though I am far from confident that I have avoided mistakes.

I would add to that sometimes such hyperbolic uses could result (1) in partial or total change of the meaning of a crucial concept or a cluster of concepts, (2) in unwilling distortions, and (3) in diverting readers in other directions. I certainly do not mean to not use the symbolic logic and analytic techniques to elucidate concepts and stated views; I rather agree with Mates and Santas, and I add some further comments, putting more emphasis on uses and misuses (Santas 1979, p. 314 n.). I am heartened to read Santas' confession and healthy skepticism in the last lines of the above passage. On this matter, P. Shorey's comments years ago seem to echo the same view: “. . . it would be irrelevant to bring in modern denunciations of the ‘old faculty psychology’” (*Rep.* 477c, p. 523, Loeb Classical Library, Vol. I). I doubt whether such practices do justice to the ancient philosophical writings, and whether they really contribute to the success of our scholarly inquiries. It seems to me, that many of our puzzlements are due in part to such practices.

(viii) I always hesitate to accept as panacea the use of modern “utilitarian” criteria or modern prevailing theories by which a critic rejects ancient theories as anachronistic and useless. For example, if a critic's thinking is dominated by such kinds of criteria, then he might conclude, along with B. Russell, that “cause is a relic of the bygone age,” doubting, disregarding or rejecting any inquiry in such old theories. He is likely to forget that the concept of *cause* is still used in medical practices and



in our everyday transactions and discussions as one of the main supporting tool in explanations and activities.

Ending my brief voyage, purporting to approximate the Platonic first and second voyages, and by way of an epilogue, I want to emphasize the following: First, I attempted to see if the concept of *aitia* is applied in each one of the ontological levels, keeping in mind (1) that Plato has not given a definition of cause, and (2) that without presupposing a definition of cause, it is difficult, if not impossible, to see whether and how the term *aitia* is used to explain objects and events in their contexts, to draw some conclusions about its nature, and to offer an account of it—even if only a tentative one. Second, I tried to develop some thoughts about: (1) the *Agathon* as cause of all (*pan-aitia*), pointing out at the same time Plato’s “arbitrary jumping” to the analogy of the Sun; (2) the key concept of *methexis* as an *ex-machina* solution; (3) the concepts of *diagnosis* and *prognosis* in terms of the Platonic aitiological framework; and (4) some questions that still remain unanswered.

**Acknowledgments** I am grateful to Gerasimos Santas and Georgios Anagnostopoulos for critical comments that helped me to clarify and correct some points in my discussion.

## Notes

1. This scheme  $A \rightarrow E$ , it seems to me, can be reversed to  $E \rightarrow A$ , and it might then lead to both a circularity and to an infinite regress.
2. This is, of course, a pseudo-causal connection (echo-shadow), based on temporal coincidence and successive regularity.
3. Plato says that seeing Simmias’ picture, he (Socrates) is reminded of Simmias; this is a clear exemplification of “the law of similarity.” He also exemplified the other, so called “laws of association,” which were systematically studied and formulated by Aristotle: similarity (*homoiois*), contrariety (*enantiois*), and proximity (*eggys*) both with respect to time and place. Aristotle’s discussion of the three “laws” of association can be found in his *Peri mnêmês kai anamnêseôs* (451b19–20). The history of use of these “laws” is very long: Hobbes, Hartley, Brown, Mill, Locke, Berkeley, and Hume. The causal connection of echo and shadow can also remind one of some views by Pavlov; the similarity has to do with rewards that reinforce such activities as prognostication, alertness and expectation of food or other rewards. The terms “prizes”, “honours” and “commendations” appear in P. Shorey’s translation of the *Rep.* (Shorey 1935), corresponding to the Greek *geras*, *timai*, and *epainoi* (*timai de kai epainoi. . . kai gerata tô(i) oxu ta horônti ta paronta*; 516c7).
4. See also *Rep.* 520c: Habituation as a process can be regarded as similar to modern *conditioning*. Plato says: “Down you must go then, each in his turn, to the habitation of the others and *accustom* yourselves to the observation of the obscure things there. For once *habituated* you will discern them infinitely better. . . and you will know what each of the ‘idols’ is and whereof it is a semblance. . .” Plato refers to those who should govern people, not those who govern us today and who “fight one another for shadows and wrangle for office. . .”
5. Also, in *Grg.* 501a: “relying on routine and *habitude* for merely preserving a *memory* of what is wont to result.”
6. *Rep.* 510a: “As the second section assume that of which this is a likeness or an image, that is, the animals about us and all plants and the whole class of objects made by man.”
7. The core of the topic is contained in the *Phd.* 95a passage about causes and causal explanation, the interpretation of which bears heavily upon Plato’s entire philosophy, at least that developed in the dialogues of the middle period. The untamable material is of immense importance for the



- understanding of the concepts pertinent to his causal theory. One should bear in mind that from 95e–107a1 of the *Phd.* the term *aitia* appears twenty times and the term *aition* seven times.
8. Though not by name, Plato seems to be referring to Alcmeon's doctrine (see Andriopoulos 1974, p. 12; 1988, p. 95; and 1971, pp. 1–7).
  9. In my judgment, Plato meant a devaluation and not a total rejection.
  10. Plato's rejection of Anaxagoras's *Nous* (though he adopted the latter's term) might be due to his failure to understand it in the way the Clazomenian thinker did, i.e., as a general, abstract concept, and not as an entity of a noetic nature, functioning as an *orderer* and *systematizer* of the many and disorderly sensible objects (see Andriopoulos 1995, 4th edition).
  11. *Rep.* 527d: "...we set down astronomy... for quickness of perception about the *seasons* and the courses of the *months* and the *years* is serviceable, not only to *agriculture* and *navigation*, but still more to the military art." Shorey's rendering of *euaisthêtotoseros* as "quickness of perception" does not seem to capture the meaning of Plato's text; "better awareness" seems more apt.
  12. *Ti.* 67b: "The third organ of perception within us which we have to describe in our survey is that of hearing, and the *causes* whereby its affections are produced. In general, then let us lay it down that *sound* is *stroke transmitted* through the *ears*, by the *action* of the *air* upon the *brain* and the *blood*, and reaching to the *soul*; and that the motion *caused* thereby, which begins in the *head* and ends about the seat of the *liver*, is *hearing*."
  13. *Rep.* 523e: "And are not the other senses also defective in their report of such things [*kai ai allai aisthêseis ar' ouk endeôs ta toiauta dêlousin*]?"
  14. *Rep.* 510c–e. See also 523e for some basic properties of the sensible objects contained in Level B: thickness and thinness; softness and hardness; light and heavy; bigness and smallness; white and black.
  15. *Phd.* 99d1–2. J.E. Raven (1965, pp. 87–8): "Plato divides Socrates' philosophical development into three stages... I am tempted to believe that Plato included the first two stages [Milesians—Empedocles, Anaxagoras] of the biography [of Socrates] partly for the purpose of understanding the supreme importance of Socrates' contribution to philosophy in contrast to that of his predecessors... he may have wished to stress those stages in Socrates' life which would throw most light on the real significance of his achievement."
  16. I.M. Crombie (1963, p. 169) concludes that Plato's thinking is rather unclear ("a nest of confusion") because (1) he lacks a semantic sensitivity and cannot elucidate the meanings of key words or phrases, and (2) he jumbles together "mathematical and non-mathematical topics."
  17. It should be mentioned that Plato's intention in *Phd.* is debated by scholars such as: J.P. Anton, (1968, pp. 94–102): "...the ideal of the good life"; R. Hackforth (1950, pp. 42–5): not to establish "a metaphysical doctrine"; Archer-Hind (1881, pp. 120–31): "...not to express purpose nor the most important philosophical result"; W.K.C. Guthrie (1981, pp. 363–4): "...the *Phd.* is about the immortality of the soul, and the posthumous blessedness of the wise and good. The doctrine of the eternal Forms itself... and the theory of recollection are ancillary to this..."; J.E. Raven (1965, p. 85): "The main topic of the *Phd.* is the divinity and immortality of the soul."
  18. At this point we must refer to the Simile of the Cave in the *Rep.* 517b, 517d, 518a: the Sun is the *aitia* of the light and vision; the light makes the faculty of sight see best and visible things to be seen (508a).
  19. Here are some of his diversified uses: *Phd.* 100e, 93e, 101c, 64e, 114c; *Rep.* 411d, 396e, 402d, 424, 519d, 455d, 455c (will it suffice us if we approximate it as nearly as possible and *metechei* of it more than others); 472c, 465ed, 432b, 525b. 510a–b (as is the opinable to the knowable so is the likeness to that it is a likeness of); 592a, 603e.
  20. Causal connections are used in an equivalent way by the above terms. See G. Santas, in Fine (1999, pp. 256, 267, 269, 265, 270); C. Shields (paper-draft): "causally responsible," "responsible for being... responsibility of the Form of Good" (pp. 8, 9, 10, 11, and several others).

21. Santas (1980, pp. 269–70). Commenting on the first draft of this paper, Santas wrote: The other Forms are “the same in all respects,” in the sense that they do not change with respect to any attributes (proper or ideal) that they have. The goodness of the other Forms is partial in the sense that it is goodness of kind, whereas the Form of the Good is good, period. The causality between other Forms and the Form of the Good is formal, but the Good and the other Forms can be final causes of human activity. But many puzzles remain, because the topic is most difficult and Plato is not explicit.
22. Santas, in Fine (1999, pp. 268–73). If the exception is convincing, then the unity (*eniaiotêta*) in Plato’s ontological structure is destroyed.
23. I am thinking that at times games (chess-like) are played in some scholarly works on the bases of verification or falsifications (supporting or rejecting) by referring to (1) concepts scattered in various dialogues (early-middle-late), and (2) evidence or non-evidence (implicit or explicit).
24. The necessity of using definitions to elucidate concepts and their functions is indispensable. Do I need to repeat here the commonplace that the search (*zêtêsis*) for definitions in almost all Platonic dialogues is both beyond doubt and a force leading upwards, step by step, to the authority of *logos*?
25. M. Frede (1980, p. 125). Critically reviewing Hume, Kant, Aristotle, Epicurus, Sextus, Chrysippus, Seneca and “what people got to think about a cause,” Frede concludes that generally cause is an active process. Concerning Aristotle, such a wide definition fits only the Stagirite’s moving cause.
26. F.-G. Hermann (2003, pp. 19–56) traces the historical usages of *metechein* in Homer, Parmenides, Zeno, the tragedians, Pindar, and Anaxagoras, showing the semantic fluctuations of the term. And for Plato, these were an accumulated, rich conceptual reservoir. In particular, the author concludes that Plato utilized the Anaxagorean terminology as well as the Anaxagorean model and applied them not to a corporeal world but to a mixed one, that is static-noetic (ideas) and dynamic-sensible (objects, events). He also offers an Anaxagorean reading of the *Phaedo*, which produces confusion and makes this dialogue problematic. This author points out the debt of Plato to his predecessors, Pre-Socratics, and Socrates.
27. R. Bolton (1988, p. 92). Bolton also suggests that Plato in the *Phaedo* “is not doing physics at all. . . [but] rejects a science of nature preparing the ground for a genuine science. . . of metaphysics.”
28. *Republic* 517b: “This [Form of the *Agathon*] is indeed the *cause* of all that is right and beautiful.” Indeed, not only of what is “right and beautiful” but, by extension, it is the cause of everything.
29. *Republic* 506b–509c: “Let us dismiss for the time being” was his answer to the question; instead, he preferred to discuss the Good’s offspring (*Agathou ekgonon*), which is similar to it. “Jumping” to the analogy of the Sun is not a satisfactory answer; nor can the “father-son” metaphor (*Hippias Major* 297b) explain the *Agathon*’s nature. I prefer characterizing Plato’s move from the *Agathon* to the Sun as a case of “jumping.” To me, the interplay between the *Agathon* and the Sun—the super-concept of the *Agathon* in the noetic realm and the Sun in the visible (*horaton*) realm—the back and forth mutual references between these two at 507b–e and 508a–509d, cannot be considered as a preparatory phase. The analogy, certainly, has a lofty poetical *aura*, but does not carry much semantic weight (ontological or epistemological) at the intelligible Levels; the nature of the *Agathon* still remains, purposefully or not, an unanswered question. It is puzzling to invoke an analogy derived from the visible (*horaton*) realm in order to prepare and explain the nature of the super-concept of the intelligible (*noê-ton*) realm. The Sun in the visible realm is causally defined by the *Agathon*. But in the analogy an inferior, sensible *x* is expected in a preparatory sense to lead to and help to define a superior, noetic *y* while, in turn, the *y* causally determines *x*. This seems circular and arbitrary—an arbitrary jumping.
30. *Rep.* 506c–e2. Commenting on the present work, Santas claims that Socrates says only that he does not *know* the Good, not that he *cannot* know the Good. But what I am saying is that he cannot *define* the essence of the *Agathon* since he does not *know* its essence. I take *legein* to

refer to the linguistic articulation of the total sum of the properties of *Agathon*. I am thinking of definition in the double sense (the identification of properties *and* their linguistic articulation), and not merely of the potential (*dynamei*) acquisition of knowledge of *Agathon* in the future. 506e can be interpreted as a temporary postponement of formulating a definition, but to my knowledge the postponement remains permanent.

31. *Republic* 509c: "The Sun. . . is the cause of the light and vision, and light and vision are sun-like but not identical with the Sun. The Good [likewise] is the cause of truth and knowledge, and truth and knowledge are like the Form of the Good, but not identical with it."
32. C. Shields (1999, p. 71; also p. 51): Holding such a view does not commit the polemical Platonist to the manifestly *false* view that *there is one Form answering to every general term*. The Platonist can certainly recognize ambiguity as well as anyone else. *Rather*, the claim now introduced is the more restricted thesis that there is only one property, Goodness, and that everything which counts as good does so because of its participating in this Form. Aristotle is sometimes credited with showing that Plato is wrong to hold that there is one Form corresponding to every general term. Surely Plato never held such a transparently false view.
33. Although saying what an X is not might help to limit the area of inquiry, it is advised by rules of constructing definitions to avoid it; it might lead into semantic impasses of the following type: baldness is not having hair.
34. G. Santas (1980, p. 381) raises the question (which is my main concern here) as to what is the sense of "cause" in Plato's theory; his quotation marks in the term *cause* indicate his hesitation in drawing a final verdict. I would be happy if he had made a distinction, or perhaps noted, the "jumping" from the conceptual framework of the Divided Line into the Simile of the Sun, as he subsequently discusses whether the Sun (*Agathon*) can be an efficient and/or formal cause of sensible things.
35. The *kath' hekasta*, especially the objects and events of the sensible world, must instantiate the characteristics included in the definitions; the same can also be said about the noetic objects, mathematical concepts and *eidê*, provided that *eidê* are taken in relation of the *Agathon* as *kath' hekasta* (and despite their status and function in relation with the sensible things). It seems to be a rather ontologically and cognitively-motivated structure explaining Plato's thought on the nature of the *Agathon* and its function as *pan-aitia*. A logical approach in explaining these metaphysical relations and interrelations by using essential definitions is preferable; there can be, if any, fewer gaps and fewer internal contradictions and inconsistencies as compared to other approaches. Further, taking this approach and using Aristotelian terms, one can take the *Agathon* as final *aitia* or *pan-aitia*, and avoid making it function as an efficient and productive cause; that is, it is viewed only as a static concept, applied to the static world of *einai*, and even to the world of *gignesthai*. Interesting is Santas' distinction between aitiological and diagnostic uses of Socratic definitions. For details see his *Socrates* (1979, pp. 97–135) where he follows A. Pap's classification of definitions.
36. G. Vlastos's view reminds one of the Natorpian interpretation of Plato's Ideas, considering them from the point of view of the laws of nature (which are Galileo's and Newton's foundations for their scientific works); Natorp contends that a similarity exists between Plato's theory and modern sciences; the *ontos on*—that is the idea—is a static concept, (a foundation) underlying the dynamic phenomena; the static concepts are not man-made, but *discovered* like the laws of nature (see *ontôs on* in *Philebus* 59d and *Sophist* 248a).
37. *Rep.* 509b2–10: "[The] objects of knowledge not only *receive* from the *presence* of the good their being known, but their very *existence* and *essence* is derived to them from it, though the good itself is *not* essence, but still transcends essence in dignity and surpassing power" (Shorey transl.; my emphasis).

## Bibliography

- Ackrill, J.L. 1997. *Essays in Plato and Aristotle*. Oxford: Clarendon Press.
- Andriopoulos, D.Z. 1971. "Alcmeon Re-examined." *Studia Classica* XIII:1–7.
- Andriopoulos, D.Z. 1974/1994. *Sense and Perception in Greek Philosophy*. New York: W. Green/Athens.
- Andriopoulos, D.Z. 1988a. *The Concept of Causality in Presocratic Philosophy*. Thessaloniki: Vanias.
- Andriopoulos, D.Z. 1988b. "The Concept of Causality in Plato." *Studia Classica*, IIX:31–8.
- Andriopoulos, D.Z. 1989. "Greek Sceptics On Causality." *Discorsi* IX:7–26.
- Andriopoulos, D.Z. 1989/1990. "Alcmeon's and Hippocrates's Concept of Aitia." *Medical Ethics* 9–15. Reprinted in Nicolacopoulos, P. ed., *Greek Studies in the Philosophy and History of Science*. Dordrecht: Kluwer Academic Publishers.
- Andriopoulos, D.Z. 1995. *Ancient Greek Theory of Knowledge*. Athens: Hestia.
- Angel, R.B. 1967. "Explanation and Prediction: A Plea for Reasons." *Philosophy of Science* 34:276–82.
- Annas, J. 1982. "Aristotle on Insufficient Causes." *Philosophical Quarterly* 32:311–26.
- Anton, J.P. 1968. "The Ultimate Theme of the *Phaedo*." *Arethusa* I:94–102.
- Anton, J.P., ed. 1980. *Science and the Sciences in Plato*. Delmar, NY: Caravan Books.
- Archer-Hind, R.D. 1881. "On Some Difficulties in the Platonic Psychology." *Journal of Philology* 10:120–31.
- Avgelis, N. 1971. *The Concept of Causality in Logical Positivism*. Thessaloniki.
- Axelos, C. 1956. "Das Einsehen und das Voraussehen im Platonischen Höhlengleichnis." *Plato* I:95–98.
- Black, M. 1955. "Why Cannot an Effect Precede Its Cause." *Analysis* 16:49–58.
- Bolton, R. 1998. "Plato's Discovery of Metaphysics: The New Methods of the *Phaedo*." In *Method in Ancient Philosophy*, edited by J. Gentzler, 91–111. Oxford: Oxford University Press.
- Braithwaite, R.B. 1953. *Scientific Explanation*. Cambridge: Cambridge University Press.
- Broad, C.D. 1935. "Mechanical and Teleological Causation." *Proceedings of Aristotelian Society* 14 (Supplementary):83–112.
- Burks, A.W. 1951. "The Logic of Causal Propositions." *Mind* 60:363–82.
- Cohen, L., and J.C. Drabkin. 1948. *A Source Book in Greek Science*. New York: McGraw-Hill.
- Cornford, F.M. 1942. "Was the Ionian Philosophy Scientific." *Journal of Hellenic Studies* 62:1–7.
- Crombie, I.M. 1963. *Examination of Plato's Doctrines* (Vol. I). London: Routledge & Kegan Paul.
- De Vogel, C.J. 1963. *Greek Philosophy I*. Leiden: E.J. Brill.
- Diels, H. 1897. *Doxographi Graeci*. Berlin: W. de Gruyter.
- Dreyer, J.L. 1953. *A History of Astronomy from Thales to Kepler* (2nd ed). New York: Dover Publications.
- Ducasse, C. 1957. "On the Analysis of Causality." *Journal of Philosophy* 54:422–26.
- Ducasse, C. 1965. "Concerning the Uniformity of Causality." *Philosophy and Phenomenological Research* 26:97–101.
- Fain, H. 1963. "Some Problems of Causal Explanation." *Mind* 72:519–32.
- Fine, G. 1987. "Forms as Causes: Plato and Aristotle." In *Mathematics and Metaphysics in Aristotle*, edited by A. Graeser, 69–112. Bern: P. Haupt.
- Fine, G., ed. 1999. *Plato I*. Oxford: Oxford University Press.
- Flew, A. 1954. "Can an Effect Precede its Cause?" *Proceedings of Aristotelian Society* 28 (Supplementary):45–62.
- Frede, M. 1980. "The Original Notion of Cause." In *Doubt and Dogmatism*, edited by J. Barnes. Oxford: Oxford University Press.
- Gallop, D. 1975. *Plato's Phaedo*. Oxford: Clarendon Press.
- Gaskin, D. 1955. "Causation and Recipes." *Mind* 54:479–87.
- Geach, P. 1956. "The Third Man Again." *Philosophical Review* 65:72–82.
- Gentzler, J. 1988. *Method in Ancient Philosophy*. Oxford: Clarendon Press.

- Georgoulis, C. 1975. *History of Greek Philosophy*. Athens: Papademas.
- Gosselin, M. 1993. "Causality: A Historical Sketch of the Major Philosophical Approaches to the Problem." *Communication and Cognition* 26:131–70.
- Grote, G. 1888. *Plato and Other Companions of Socrates* (4th ed.). London: J. Murray.
- Guthrie, W.K.C. 1967. *A History of Greek Philosophy* 1. Cambridge: Cambridge University Press.
- Guthrie, W.K.C. 1981. *A History of Greek Philosophy* 2. Cambridge: Cambridge University Press.
- Hackforth, R. 1950. "Immortality In Plato's *Symposium*." *The Classical Review* 64:43–45.
- Hackforth, R., trans. 1955. *Plato's Phaedo*. Cambridge: Cambridge University Press.
- Hackforth, R. 1972. *Plato's Phaedo*. Cambridge: Cambridge University Press.
- Hermann, F.G. 2003. "Metetechein, metalambanein and the Problem of Participation in Plato's Ontology." *Philosophical Inquiry* XXV(3, 4):19–56.
- Hintikka, J. 1973. "Knowledge and Its Objects in Plato." In *Patterns in Plato's Thought*, edited by J.M.E. Moravcsik, 31–51. Dordrecht: Reidel.
- Houssiadis, L. 1964. *An Exploratory Study of the Perception of Causality*. Cambridge: Cambridge University Press.
- Hutten, E.H. 1962. *The Origins of Science*. London: George Allen & Unwin.
- Irwin, T. 1977. *Plato's Moral Theory: The Early and Middle Dialogues*. Oxford: Oxford University Press.
- Kahn, C. 1996. *Plato and the Socratic Dialogues*. Cambridge: Cambridge University Press.
- Keyt, D. 1969. "Plato's Paradox that the Immutable is Unknowable." *Philosophical Quarterly* 19:1–14.
- Korsgaard, C. 1983. "Two Distinctions in Goodness." *Philosophical Review* 92(2):169–95.
- Kraut, R., ed. 1992. *The Cambridge Companion to Plato*. Cambridge: Cambridge University Press.
- Lloyd, A.C. 1976. "The Principle that the Cause is Greater than Its Effect." *Phronesis* 21:146–155.
- Mackie, J.L. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2:245–64.
- Malcolm, J. 1991. *Plato on Self-predication of Forms*. Oxford: Oxford University Press.
- Mathews, G.B., and T.A. Blackson. 1989. "Causes in the *Phaedo*." *Synthese* 79:581–91.
- Michotte, A. 1954. *La Perception de la Causalité*. Louvain: Publications Universitaires.
- Nehamas, A. 1972–1973. "Predication and Forms of Opposites in the *Phaedo*." *Review of Metaphysics* 26:461–91.
- Nehamas, A. 1975. "Plato on the Imperfection of the Sensible World." *American Philosophical Quarterly* 12:105–17.
- Nussbaum, M. 1986. *The Fragility of Goodness*. Cambridge: Cambridge University Press.
- O'Brien, D. 1967. "The Last Argument of Plato's *Phaedo* I." *Classical Quarterly* 17:198–231.
- Owen, G.E.L. 1968. "Dialectic and Eristic in the Treatment of Forms." In his *Aristotle On Dialectic*, 103–25. Oxford: Clarendon Press.
- Pappas, N. 2006. *Plato and the Republic*. London: Routledge.
- Penner, T., and C.J. Rowe. 1994. "Desire for the Good." *Phronesis* 39(1):1–25.
- Pfeiffer, R. 1968. *A History of Classical Scholarship I*. Oxford: Oxford University Press.
- Popper, K. 1966. *The Open Society and Its Enemies*. Princeton: Princeton University Press.
- Raven, J.E. 1965. *Plato's Thought in the Making*. Cambridge: Cambridge University Press.
- Rawls, J. 1971. *A Theory of Justice*. Cambridge: The Belknap Press.
- Robinson, R. 1953. *Plato's Earlier Dialectic*. Oxford: Oxford University Press.
- Rose, L. 1966. "The Deuterous Plous in Plato's *Phaedo*." *The Monist* 50:464–73.
- Ross, W.D. 1951. *Plato's Theory of Ideas*. Oxford: Oxford University Press.
- Russell, B. 1912–1913. "On the Notion of Cause." *Proceedings of the Aristotelian Society* 12:1–26.
- Santas, G. 1979. *Socrates*. London: Routledge & Kegan Paul.
- Santas, G. 1980. "The Form of the Good in Plato's Republic." *Philosophical Inquiry* 1:374–403.
- Santas, G. 1985. "Two Theories of Good in Plato's Republic." *Archiv für Geschichte der Philosophie* 57:223–45.
- Santas, G. 1986. "Plato on Goodness and Rationality." *Revue Internationale de Philosophie* 40: 97–114.
- Santas, G. 1988. *Plato and Freud*. Oxford: Blackwell.

- Santas, G. 1993. "Does Aristotle Have a Virtue Ethics?" *Philosophical Inquiry*, 3–4:1–33.
- Santas, G. 2001. *Goodness and Justice: Plato, Aristotle, and the Moderns*. Oxford: Blackwell.
- Santas, G. 2006. "Justice, Law, and Women in Plato's *Republic*." *Philosophical Inquiry*, XXII: 91–103.
- Sarton, G. 1960. *A History of Science*. Cambridge: Harvard University Press.
- Sayre, K.M. 1969. *Plato's Analytic Method*. University of Chicago Press.
- Shields, C. 1999. *Order in Multiplicity*. Oxford: Oxford University Press.
- Shorey, P. 1935. *Plato: The Republic*, Vol. I & II. Cambridge: Harvard University Press.
- Shorey, P. 1948. "The Idea of the Good in Plato's *Republic*." *University of Chicago Studies in Classical Philology*, 1:188–239.
- Taylor, A.E. 1928. *A Commentary On Plato's Timaeus*. Oxford: Clarendon Press.
- Van der Waerden, B.L. 1961. *Science Awakening*. Groningen: P. Noordhoff.
- Vanquickenborne, M. 1969. "An Analysis of Causality in Everyday Language." *Logique et Analyse*, 12:311–28.
- Virvidakis, S. 1989. "A Metaethical Approach to Plato and Aristotle." *Philosophical Inquiry* 11: 38–64.
- Vlastos, G. 1967. "Degrees of Reality in Plato." In *New Essays in Plato and Aristotle*, edited by R. Bambrough, 219–34. New York: Humanities Press.
- Vlastos, G. 1969. "Reasons and Causes in the *Phaedo*." *Philosophical Review*, 78:291–325.
- Vlastos, G. 1981. *Platonic Studies*, 187–96. Princeton: Princeton University Press.
- Vlastos, G. 1987. "Separation in Plato." *Oxford Studies in Ancient Philosophy* 5.
- Vlastos, G. 1991. *Socrates, Ironist and Moral Philosopher*. Ithaca: Cornell University Press.
- Wagner, M.F. 1979. *Concepts and Causes. The Structure of Plotinus' Universe*. PhD Dissertation. Ohio State University.
- Windelband, W. 1958. *A History of Philosophy I*. New York: Harper.
- Wright, G.H. von 1966. *The Varieties of Goodness*. New York: Humanities Press.
- Yonezawa, S. 1991. "Are the Forms *aitiai* in the *Phaedo*?" *Hermes* 119:37–42.
- Yu, J. 1991. "On Plato's Theory of the *Methexis* of Ideas." *Philosophical Inquiry* 25–38.
- Zeller, E. 1922. *Outlines of the History of Greek Philosophy*. London: Longmans, Green, and Co.

# Chapter 19

## Desires and Faculties in Plato and Aristotle

Deborah K.W. Modrak

In *Goodness and Justice: Plato, Aristotle, and the Moderns*, Gerasimos Santas offers a wonderfully comprehensive interpretation of Platonic and Aristotelian theories of the good.<sup>1</sup> Central to this interpretation is the distinction between desire satisfaction theories of the good and functional/perfectionist theories of the good. Plato rejects the former and defends the latter; Aristotle subscribes to both. Santas' emphasis on the difference between the two types of theories of the good and the way these are dealt with in Plato's and Aristotle's accounts of the human good has many implications for the interpretation of Platonic and Aristotelian philosophy. In this paper, I shall look at only one of the many subjects Santas addresses—Plato's and Aristotle's conception of desire.

### 19.1 Platonic Desires

Plato has many things to say about desire in a variety of texts. Santas discusses the *Gorgias*' and *Republic*'s views in detail. Socrates' argument with Polus in the *Gorgias* shows that "our wanting things as means is constrained by truth about the relation of the means to our good" (Santas 2001, p. 26). The issue is whether despots do what they want to do when they exercise unrestrained power (466b–468d). Socrates denies that they do, and ultimately gets Polus to concede this point. While Santas' analysis of Socrates' argument is right on track, we might still worry that Socrates/Plato is confusing two things—desire as actually experienced and an idealized or normative notion of desire in terms of what should be desired. What it is rational to desire need not be what in fact is desired. Perhaps Polus should have replied, "I agree that they would attempt to reform their desires were they aware that these desires are not, all things considered, in their best interest, but that does not call into question their having the desires they have right now." Indeed this view of desire seems to be assumed by both Callicles and Socrates in their disagreement

---

D.K.W. Modrak (✉)  
University of Rochester, Rochester, NY, USA  
e-mail: deborah.modrak@rochester.edu



about whether the life of pleasure is the best life (481c ff.). Socrates likens the life spent in the pursuit of pleasure to constantly refilling a leaky jar. Callicles retorts that pleasure is located not in the static state of having no desires but in the dynamic state of satisfying occurrent desires (494a). Socrates responds that then pleasure and pain must co-exist and cease at the same moment (497c). Ultimately Callicles grants that some pleasures are better than others and this paves the way for Socrates to conclude that one should restrain one's desires in the interest of achieving the best life possible (505b). Here the life of pleasure is identified with the life of desire satisfaction and desires are taken to be impulses towards various objects that may or may not be good for us. The paradigm case of a desire is that of a basic bodily impulse such as hunger or thirst. Since such desires are need driven, Socrates' talk of replenishment and refilling seems appropriate, as does the attribution of a narrow concept of desire as bodily appetite to Plato.

Plato's initial discussion of psychic partition in the *Republic* suggests a similar picture. The contours of the lowest part of the soul are determined by appeal to a notion of desire that seems quite like the one found in the *Gorgias*. Many commentators have emphasized the role of psychic conflict in the partition of the soul in *Republic* IV. Santas, by contrast, emphasizes the role of the characteristic objects and exclusive functions of the three basic types of psychological activities, viz. desiring, getting angry and calculating in grounding the separation of psychic faculties. "I believe that these two ideas, of exclusive functions and characteristic objects, are the main principles at work for individuating the parts of the soul; the partition itself, as based on psychic conflict and the principle of contrariety, is not sufficient for understanding the nature of each part" (2001, p. 122). This is a powerful insight and one that brings Plato's psychological theory much closer to that of Aristotle than is commonly acknowledged:

In the case of a faculty I look to that alone which it is related to and what it accomplishes and in this way I call each of them a faculty and that which is related to the same thing and accomplishes the same thing I call the same faculty and a faculty which is related to and accomplishes another thing, I call other. (*Rep.* V 477 cd)

When Aristotle adopts this approach to the differentiation of faculties, it results in a proliferation of faculties. We might well wonder why in Plato's hands this strategy yields only three parts of the soul and we might also wonder whether this outcome is consistent with his methodology. With respect to desire, Aristotle argues against partition on the grounds that desire (*orexis*) includes bodily desire (*epithumia*), spirited desire (*thumos*) and rational desire (*boulêsis*) (*An* III.8 432b3–8; III.10 433a21–26), and thus desire cannot be limited to a single part of the soul. In *Republic* IX, Plato mentions desires that are proper to spirit and reason, but arguably this threatens the differentiation of the faculty of desire in terms of its characteristic activity of desiring (580d). One way that Plato could save his account would be by positing a unique object that differentiates the exercise of the desiderative faculty from any other type of desire. Whether he does or not remains to be seen.

In *Republic* IV, having identified the three basic psychic activities, Plato makes each a function of a distinct faculty. Basic desires such as hunger and thirst have

unique objects, Socrates argues (437a–439a). The object of thirst is drink; of hunger, food. He resists any attempt to qualify these generic objects as good or bad, tasty or bitter, etc. When one is thirsty, her soul desires nothing other than drink. However, the agent may choose not to drink, because something else in the soul forbids drinking. This observation grounds the separation of the desiderative (*epithumetikon*) and irrational part (*alogiston*) of the soul from the rational part (*logistikon*) of the soul (439d). Socrates next argues for the distinctness of the spirited part (*thumoeides*) from the rational part (441a–c). When our desires conflict with our judgment, we experience opposing motivations. Sometimes spirit sides with desire against reason, and this would not happen were spirit not distinct from reason. In addition, reason and spirit do not always co-exist. Children and beasts possess a capacity for anger but not for reason. This argument establishes distinct functions for the three psychic faculties but leaves their objects undefined.

If there is a distinctive generic object that defines either of the higher faculties, there is no mention of it. Nor is the argument from object to faculty completely clear in the case of the desiderative faculty. The objects mentioned are peculiar to specific desires and it remains an open question whether they could all be collected under some more generic kind. The faculty itself is described as a companion of replenishments and pleasures (439d). The mention of replenishment seems to be a reference to the named desires for sex, food and drink, but the mention of pleasures is much more open-ended, as is Socrates' inclusion of "the other desires." Perhaps, Plato intends "pleasure" to be the generic object of the desiderative faculty—at least in *Republic IV*.<sup>2</sup> This seems all the more likely in light of the *Gorgias*' association of pleasure and desire.

Socrates claims that the desiderative faculty is the greatest part of the soul (*pleiston*) (*Rep. IV* 442a). Its sheer bulk together with bodily replenishments and pleasures threatens to overwhelm the other parts of the soul, which must be vigilant in keeping it in check. In the just individual, each part of the soul performs its proper function. But precisely what this function is in the case of the desiderative part is not spelled out. Presumably, it is to ensure the physical well-being of the individual in so far as this is consistent with the aims of the other psychic parts. This suggests a narrow construal of the desiderative faculty as the capacity for various bodily appetites, and many commentators, including Santas, have adopted this reading. "The natural, inborn, unlearned power of appetite operates on the generic objects of food, drink, and sex; these are its most evident objects; we understand what this power is by reference to these objects and the activities which satisfy or extinguish the appetites" (Santas 2001, p. 123). However, Plato may not intend in *Republic IV* to restrict the desiderative part only to bodily appetites. The mastery of the soul by the desiderative part is associated with a person being able to do whatever he wishes (*boulêthê(i)*) except what would free him of badness and injustice (445b). This would seem to be a broader conception of desire than that of bodily appetites.

In *Republic VIII* and *IX*, as Santas points out, not only the desiderative faculty but also the other two basic psychic faculties have desires (120–25). In *Republic VIII*,

the spirited person is described as a lover of victory (*philonikos*) (550b). In *Republic* IX, Socrates argues that each faculty has a characteristic desire and pleasure:

It seems to me that there are three pleasures corresponding to the three parts of the soul, one peculiar to each part, and similarly with desires and kinds of rule. (*Rep.* IX 580d)

By not confining all desiring to the desiderative part, Plato has greatly complicated his account of desire and arguably has made it incoherent. In *Republic* IV, Socrates argues from the basic activity of desiring to the division of the soul into a desiderative part and a rational one. If desiring is common to all psychic faculties, then desiring can no longer be cited as a distinctive activity that defines a fundamental faculty. Plato nowhere addresses this difficulty. Moreover, the account of desire in *Republic* VIII and IX is further complicated by Plato's need to distinguish between three types of individuals whose souls are dominated by the desiderative faculty, viz. the oligarch, the democrat and the tyrant.

Oligarchy is associated with the lowest part of the soul (*Rep.* VIII 551a–554a). The oligarch, however, is described as a lover of things (*philochrêmatos*). This suggests a broader conception of the desiderative faculty than that of a faculty limited to bodily appetites. The oligarch keeps his other bad desires under control by a suitable force (*epieikei bia(i)*) and for the most part his better desires conquer his worst ones (554c–e, 558d). There is some ambiguity here. Does the force that holds these desires in check originate in the spirited or rational part of the soul or is it simply a manifestation of the intensity of the desire for wealth that overpowers everything else? The latter interpretation gains some support from the description of the oligarch as concerned lest his desires become allies of spirit (555a). According to Santas, the oligarch's desire for wealth is "a mixture of pure appetites and learning" (2001, p. 124). Such mixed desires are not functions of the desiderative faculty alone but complex activities produced by a combination of different basic and pure psychic activities of several parts of the soul (2001, p. 124). An alternative interpretation would grant that desires are conditioned by life experiences involving other parts of the soul but would insist that conditioned desires are, nevertheless, activities belonging to the desiderative faculty. This would explain why the desire for wealth that is distinctive of the oligarch is identified as a life where the desiderative faculty is dominant.

Turning to the democrat, Socrates distinguishes between necessary and unnecessary desires, calling the former, productive of wealth (*chrêmatistikas*) and the latter wasteful (*analôtikas*) (*Rep.* VIII 558d–59c). The desire for bread is a necessary desire; the desire for meat and other such foods is unnecessary. While the oligarch is ruled by necessary desires, the democrat is ruled by unnecessary desires (559c–d) that are directed at entertainment and display (572c). The oligarch has a single dominant desire that structures and suppresses his other desires; the democrat gives all desires equal footing (560a–b; 561b). The democrat refuses to acknowledge that some pleasures originate from good and noble desires, while other pleasures originate from base desires (561c). The democrat sometimes indulges in wine and unhealthy foods and, at other times, diets and exercises. The picture of the democrat as someone who refuses to discriminate among her desires, pursuing first this one

and then that one, provides further evidence for the presence of an array of desires in the desiderative faculty. Plato speaks of the desire of the day and gives the following examples—drinking wine, listening to flute music, dieting, exercising, and even dabbling in philosophy (561c–d). Such descriptions bring the question of the unifying principle of the desiderative faculty into vivid focus. Is there, available to Plato, a generic description of the object of the faculty that distinguishes this faculty from the other two parts of the soul?

At the beginning of the sketch of the tyrannical character in *Republic IX*, Socrates says that we must look again at the nature and number of human desires (571a). In particular, he directs our attention to the unnecessary pleasures and desires that are lawless. These exist in all of us but are held in check by the better desires and reason (*logou*) (571b–c). That the better desires act in concert with reason shows that reason and desire act together on occasion. Plato's language also suggests, however, that although some desires are better than others, the better desires are not themselves expressions of reason. Socrates finds evidence for the lawless desires in our dreams and recommends that we quiet our desiderative faculty by neither depriving nor sating ourselves before sleep in order to encourage lawful, rational dreams (571c–572b). The tyrannical personality is one in which Eros has become the dominant desire that protects and nurtures all the other desires (572e–74a). When this happens, lawlessness rules in the soul and the tyrannical individual seeks to gratify Eros and his other desires without regard to other people or institutions. All the better parts of his soul are enslaved and driven to serve the worst and most frenzied part of his soul (577d). Interestingly, Plato does not liken the tyranny of Eros to the rule by the desiderative part of the soul as a whole (the analogue in the individual of the producer class) but rather to the rule by one part of it.

Because Plato needs an account of the desiderative faculty that allows him to distinguish three types of rule (oligarchy, democracy and tyranny), he ends up with a picture of this faculty as made up of a large and diverse number of desires. In order to bring unity to this collection, Plato looks for a generic description and finds it in the love of wealth (581a). This device provides a weak sort of unity to the faculty and frames the differentiation of three kinds of relations that the component desires may exhibit. If the desire for wealth is stronger than the other desires, it subordinates them, causing the oligarchic individual to be temperate, since the cultivation of bodily appetites would undermine the control of the desire for wealth. If many, diverse desires relate to one another as equals, they cause ever-shifting motivations and erratic actions but, since the better desires are sometimes in control, the democrat does not behave in a truly depraved manner. If Eros—the most lawless and primitive of desires—gains the upper hand, then internal psychic chaos and external depravity reign. The unity provided by the presence of the love of wealth in all three types of individual (oligarch, democrat and tyrant) is weak in that the desire for wealth is only one of many desires that are characteristic of the democrat and the tyrant. Plato makes no effort to show that the desire for wealth is more basic or that the other desires somehow flow from it. On the contrary, he emphasizes the way that the other desires overwhelm the desire for wealth in democratic and tyrannical individuals.

The tyrannical man is as lawless as his internal state. He is also, according to Plato, the most miserable of men. He is miserable, if he is a private citizen, because he is continually tormented by his unsatisfied desires and by the unfortunate state of his soul (*Rep.* IX 579e). He is made mad by his desires and passions (578a). He is even more miserable, if he is an actual tyrant, because in that case he not only has internal strife to deal with but also external threats to his power. The second argument Socrates offers for this conclusion (580d–83a) is directly relevant to determining the range of the desiderative faculty. Returning to the partition of the soul, he claims that each part has a pleasure and desire peculiar to it (580d). The third part of the soul has many forms but is called desiderative (*epithumetikon*) after its strongest part, the desires for food, drink, sex, and its love of wealth (580e). It is rightly brought together under the description lover of wealth (*philochrêmaton*) and lover of gain (*philokerdes*) (581a). The spirited part desires victory, and the rational part, wisdom. Here certain kinds of desires, viz. those for victory and wisdom, are excluded from the purview of the desiderative faculty. We also find a large number of non-bodily desires included among the desires of the desiderative faculty. The next stage of the argument distinguishes the three pleasures corresponding to the three types of desire and finds that the lover of wisdom's pleasure is far superior to that of the lover of victory or gain.

To wrap up his argument for the superiority of the life of the just individual, Socrates sets out to show that the pleasure experienced by the other parts of the soul is illusory in comparison to the pleasure experienced by the rational part (583b–88a). In the case of the pleasures belonging to the lower parts of the soul, pleasure is typically the cessation of pain; and pain, the cessation of pleasure. In the case of sensing, pure pleasures can be found in association with the body but most bodily pleasures are mixed with pain (584b–c). This makes their character illusory and ever-changing. Despite Plato's conclusion, it is worth noting that the argument turns on taking pleasure to be the generic object of desire and then showing that the pleasure of the rational part of the soul is superior to that of the other parts of the soul:

If, then, to be filled with what is appropriate to our nature is pleasure, then that which is more really filled with real things would more really and truly enjoy a true pleasure, while that which partakes of the less truly real is less truly and surely filled and partakes of a less trustworthy and less true pleasure. (*Rep.* IX 585d–e)

Pure pleasure is characteristic of the exercise of the rational part of the soul by the lover of wisdom. This conclusion reveals Plato's motivation for introducing desires and pleasures that belong to the two higher parts of the soul. Had the rational soul no pleasure peculiar to it, the just life, the happiest life, would be without pleasure. This would seem quite paradoxical—even to Plato. As Callicles remarks in the *Gorgias*, a pleasure that is merely the absence of desire seems rather hollow.

Unfortunately, assigning desire and pleasure to the other two faculties is in tension with the division of the soul into three parts. The division seemed straightforward when all desires and pleasures belonged to the desiderative faculty. The desire for wealth and the desire for victory are distinguishable with respect to their objects

but qualitatively they seem very similar. Neither can be reduced to a mere bodily appetite. They are uniquely human desires, etc. This may be one of the reasons why Santas emphasizes the role of learning in both cases and labels them mixed states. However, as presented in *Republic IX*, Plato distinguishes them only by reference to their objects and the faculties to which he assigns them. The satisfaction of a desire of any sort yields pleasure; the quality of the pleasure varies with the quality of the kinds of objects or activities that are desired.<sup>3</sup> In this argument, Plato appeals to the notion of bodily pleasures and pains, but as we have seen, to limit the desiderative faculty to bodily urges makes a hash of assigning all three forms of degenerative constitutions to the political analogue of the desiderative faculty.

The challenge facing Plato in the *Republic* is to give an account on which both the tripartite division of the soul based on the activities of desiring, being angry and reasoning and their characteristic objects and functions is consistent with distributing desires and pleasures across all three psychic faculties. On Santas' interpretation, the tension between these positions is resolved because the partition of the soul in *Republic IV* is a division into inborn capacities, whereas the desires that Plato describes in the later books of the *Republic* are desires mixed with reason, not mere bodily drives (2001, pp. 122–5). This is a great advantage of his reading. Yet it is perhaps worth noting that on a first reading of Plato's text the tensions within his account are more evident than is their resolution. As described by Plato, the characteristic activity of the desiderative faculty is not consistently limited to inborn appetites. As we have seen, it is not an easy matter to identify a generic object of the desiderative faculty. The two strongest candidates in *Republic IV* would seem to be bodily pleasure and pleasure more generally conceived. When we turn to *Republic VIII* and *IX*, it is clear that neither of these candidates will do. It must have also been clear to Plato, as he ultimately identifies the generic object of the lower part of the soul with wealth. In order to maintain the analogy with the kinds of states, however, Plato not only appeals to the generic object, wealth, to define the desiderative faculty but he also appeals to the relation among various objects of desire. In view of these difficulties, it may not come as a surprise that Aristotle uses desire as a case in point to argue against a tripartite division of the soul (*An III.9* 432b3–8, *III.10* 433a21–26).

## 19.2 A Better Account

Aristotle adopts the principle that Plato formulates in *Republic IV* and consistently identifies psychic faculties by means of their distinctive objects and characteristic functions. As a consequence, in his psychological writings, Aristotle offers an integrated model of psychic functioning based on distinct faculties, objects and activities.

There are three types of desire that differ by definition and capacity (*An III.10* 433b3–5). They are sensual desire (*epithumia*), emotion (*thumos*) and rational desire (*boulêsis*). Sensual desire aims at bodily pleasure and seeks to avoid bodily pain;

rational desire aims at the good and emotion falls somewhere in-between. Rational desire belongs only to humans; non-rational animals and young children possess both sensual desire and emotion (*NE* III.2 1111b12–13). Because there are different types of desire, conflicts between desires may arise. Whenever we have conflicting motivations, Aristotle argues, we have conflicting desires. Cases where an agent appears to act against her desires are, for Aristotle, cases where the agent's rational desire has conflicted with and bested the agent's sensual desires or emotional ones. They are not cases where reason as such has become a motive capacity. On Aristotle's account, when an agent acts on principle and against some of her desires, she is still acting on desire. While different types of desire may conflict with one another, potentially conflicting objects of desire falling under a single type of desire will not produce actually conflicting desires of that type. Suppose I am thirsty and hungry and objects that would satisfy both are ready to hand. To act I must pick up one or the other. If the thought of a cool drink provokes more pleasure than the thought of food, I'll drink first. The sensual desire that was fully realized had a single focus. Sensual desire will always settle on what appears to be most pleasurable or least painful (*NE* III.2 1111b16).

In the *De Anima*, Aristotle recommends an analytic strategy that works back from the object of a psychic faculty to its characteristic activity to the faculty itself. The faculty is a capacity for a particular type of psychic activity; the activity has the character it has in virtue of being directed upon a particular object. This is a method that allows Aristotle to provide as fine-grained an account of a particular psychic capacity as the context demands. He applies the same analytic technique to the characterization of a specific instance of seeing a particular shade of red here and now, to the characterization of the general activity of seeing red and to the characterization of the even more general activity of seeing color. In all three descriptions, the psychic activity is defined in terms of its object and the faculty in terms of the activity. Sight is the capacity for seeing color; this capacity encompasses the capacity to see red, which encompasses the capacity to see a particular shade of red. Similarly, the capacity for desiring food encompasses the desire for a particular type of food and this desire may be manifested in the desire for a particular delicacy here and now. Employing this line of analysis, Aristotle is able to present a generic capacity such as perception or desire as a unified whole encompassing a number of more specific capacities—for instance, sensual desire, passion and rational desire, and to treat the more specific capacities as unified despite their having numerous distinct realizations.

Desire is of interest to Aristotle, because he wants to explain the motive faculty of the soul. His question is: What gets us moving and why? Aristotle's theory of motion provides the context for his analysis of desire (*An* III.10 433b10–21). The faculty of desire is that which moves us in virtue of being a psychic activity that prompts the movement of the relevant bodily parts—muscles, sinews, and limbs and so forth—as is clear from the physiological account of desire and self-movement in *De Motu Animalium*. Desire is a moved mover and its object, the thing desired, is the unmoved mover of the desire (*An* III.10 433b11; *MA* 703a5). The object of desire prompts the desire; the desire for the object moves the agent to act. Aristotle considers and rejects the possibility that some other psychic capacity such



as perception or thought occasions action. He concludes that only desire in one of its various forms is sufficient to cause a movement that is initiated by an agent, human or non-rational animal. Aristotle defends his position by pointing out that we can entertain the thought of something pleasurable or frightening and not be motivated to act. Perceptions do not provoke actions unless desire is present.

Aristotle is, arguably, much clearer than Plato about how various forms of cognition (perception, belief, etc.) interact with our desires. The objects upon which particular desires are focused are presented to the agent by some form of cognition. Broadly speaking, this may be a perceptual cognition (perception, imagination, or memory) or a rational cognition (*An* III.12 434b12; *MA* 701a28–b1). The cognition is, on the one hand, a presentation of a particular object and, on the other, a presentation of it as pleasurable or otherwise desirable. In order to perceive a piece of moussaka as desirable, one must not only recognize that it is a piece of moussaka, but one must also perceive it under a description that makes it desirable. This description would connect the moussaka with one's desire for food, with one's liking the flavor of eggplant, and so forth. A person who has just eaten is unlikely to perceive a piece of moussaka as desirable, even though he may quite like moussaka and he recognizes that the object in front of him is moussaka. In order for an object to prompt desire and hence an action, it must present itself to the agent as actually desirable at that moment. To perceive the moussaka as desirable is to apprehend it as a good for oneself here and now. A prospective agent may believe, for instance, that one should eat when hungry and since he is currently hungry, he will eat the moussaka. Desire, like perception, is realized through the actualization of its object as an object of awareness. It differs from perception and other cognitive functions in being the realization of affective characteristics. The object of desire is a cognitive object, an object of perception or thought, embedded in a presentation that not only has cognitive but also affective qualities. In the case of humans, the latter may be articulated as universal principles. In the case of other animals, there is no capacity for articulating general principles but the object of desire is, nonetheless, presented as desirable:

Since one judgment and premise is universal and the other particular, for the one says that a person of this sort ought to do such and such, and the other says that this is such and such, and I am such a person, this opinion causes action, not the universal, or perhaps both; but the universal is more at rest; and the particular is not. (*An* III.11 434a16–21)

The particular judgment or presentation moves the agent because it occasions the presentation of an object, obtainable through action, as an object of desire. The psychic realization of an object of desire is a desiring. Actively desiring the object causes the agent to move to obtain it.

Broadly speaking, all forms of desire aim at a good that may be achieved through action. Aristotle calls this kind of good the practical good (*An* III.10 433a30). The theory of desire that is sketched by Aristotle in his psychological writings is filled out in more detail in his ethical ones. Desiring the right thing is central to Aristotle's account of virtue. Virtue is a disposition to behave in various ways. A mark of virtue is taking pleasure in the right objects. Unless our desires are in line with our moral

judgments, we will not be able to behave virtuously. Aristotle makes choice (*prohairesis*) a necessary condition for the possession of virtue, and he defines choice in terms of desire. Choice is deliberate desire (*bouleutikê orexis*) (*NE* III.3 1113a11, VI.2 1139a23). Choice is desire for an end achievable through our own action that arises after deliberation. Sensual desire figures importantly in Aristotle's accounts of temperance and its opposite, profligacy, and weakness of will (*akrasia*) and its opposite, self-restraint. Aristotle draws a distinction between natural sensual desires and ones that are peculiar to individuals (*NE* III.11 1118b8–22, VII.4 1147b24–31). Natural sensual desires, such as those for food and drink and sex, are common to all humans. Peculiar desires are often permutations of natural ones, for instance, a desire for delicacies. People are much more likely to have excessive or inappropriate peculiar desires than natural ones. Aristotle's treatment of peculiar desires is reminiscent of Plato's discussion of unnecessary desires in *Republic* VIII. Aristotle uses the same examples as Plato does; both underscore the difference between a basic need such as the desire for food and desires for specific foods. Appealing to his analysis of desire, Aristotle explicates the nature of temperance and self-restraint and the corresponding cases of morally problematic behavior. The temperate person has natural sensual desires that are appropriate to the occasion (*NE* VII.2 1146a12). The self-restrained person acts on rational desire that holds his inappropriate sensual desires in check (*NE* VII.1 1145a17, b8, VII.2 1146a10). In a broader sense, however, all virtuous action involves desiring the right objects at the right time and choosing the right means to achieve them.

This brings us to Plato's and Aristotle's handling of the desire for happiness. Both hold, and Aristotle argues, that all agents aim at happiness. The desire for happiness seems absolutely basic and yet quite different from other desires. Neither philosopher identifies happiness (*eudaimonia*) with a subjective feeling. Aristotle seriously considers the question whether anyone can be called happy while still alive. Not only is happiness an objective state that must be evaluated from a long-term perspective, it cannot be the thing that motivates a particular action. We choose, i.e., deliberately desire, ends that are achievable through specific actions (*NE* III.2 1111b20–29). We do not choose to be healthy, Aristotle says, but to eat healthily on a particular occasion. There is a form of desire that takes ultimate ends as its object. This is rational desire for ends such as health, happiness or even immortality. Under the most general description, the object of rational desire is the good, i.e., what is objectively good, or the apparent good—what the agent takes to be good (*NE* III.4 1113a23–30). The good individual construes the final end correctly; the bad individual does not. Santas says that for Plato the desire for happiness is a desire of the rational part of the soul. This seems quite likely but Plato does not say as much. Santas considers several arguments and ultimately justifies his reading on the basis of the role that pleasure plays in generating value: “So by the successful practicing of its activities reason can generate enjoying them and enjoying them can generate desires of reason for these activities, their objects and results” (2001, p. 121). Nor does Plato explicitly provide an account of the way that the desire for happiness structures the actions of individuals. The argument of the *Republic*, if successful, establishes that the just individual is truly happy but what role her desire for happiness plays in her actions is left unexplained. Here, too, arguably Aristotle's analysis

is better, because the tri-partite division of psychological faculties and activities has been jettisoned. He is able to account for immediate motivations in terms of desires that are directed upon goods that are obtainable through specific actions and at the same time to present these desires within a framework of ultimate ends that are the objects of rational desire. When Plato identifies a desire that is specific to the rational part, not surprisingly he makes it the desire for truth. The reader is left to work out the connection between the desire for truth and the desire for happiness.

### 19.3 Conclusion

Santas' analysis of psychic partitioning in the *Republic* emphasizes the relevance of Plato's conception of psychic faculties to the tri-partite division of the soul. Following this important insight, we have looked at Plato's account of desire and at Aristotle's. We have found that Aristotle accepts Plato's method of individuating psychic faculties by their objects and characteristic functions and applies it more systematically. This enables Aristotle to arrive at a more adequate account of desire than Plato's. His account is not constrained by a rigid division of the soul into desiderative, spirited and rational parts. Nor does he have to find analogues for three types of flawed states in the desiderative faculty. Aristotle recognizes three basic types of desire and three objects of desire that correspond roughly to the different types of desire posited by Plato in *Republic* IX. However, unlike Plato, Aristotle assigns all three fundamental types of desire to a faculty for desire that is itself a sub-system of the locomotive faculty of the soul. Desires of all sorts motivate us to act and this is the distinctive feature of the faculty as a whole. That said, Plato deserves a great deal of credit for providing the analysis of faculties that yields a powerful account of psychological functions in his own hands and in those of his student.

### Notes

1. All references to Santas (2001) below will be to this book unless otherwise indicated.
2. In *Republic* IX, Plato introduces pleasures and desires that are peculiar to the higher parts of the soul. The ways in which this complicates his account of desire will be discussed below.
3. This point is first made at *Rep.* VIII 561c. Plato mentions the democrat's failure to distinguish between pleasures arising from good and noble desires and pleasures arising from base desires.

### Bibliography

Santas, G. 2001. *Goodness and Justice: Plato, Aristotle and the Moderns*. Oxford: Blackwell.

# Chapter 20

## Is Aristotle's Function Argument Fallacious?

### Part 1, Groundwork: Initial Clarification of Objections

Gavin Lawrence

#### 20.1 The Problem

The most notorious problem with Aristotle's *Function Argument* in *NE* I.7 is the charge that its main conclusion—the conclusion that

the human good (*to anthrōpinon agathon*) is a <reason-involving> activity of soul in accord with its <proper> excellence<sup>1</sup>

—is the result of a fallacious inference.<sup>2</sup> But, for all its notoriety, there is a considerable lack of clarity and precision, among detractors and defenders alike, over what the alleged fallacy is.

Glassen objects that “there can be no doubt that Aristotle did confuse the notion of *the goodness of* with the notion of *the good of* man” (1957, p. 322). Ackrill objects that “it is not self-evident that the best thing *for* a man is to be the best possible man” (1973, p. 20). Wilkes objects that there is a logical gap between “the life of the good man” and “the life that is good *for* a man” (1978, pp. 343, 345)—that

if happiness is indeed the greatest good for man, excellence of functioning seems *neither to entail it nor be entailed by it*. (p. 343, my emphasis)

(Wilkes here takes for granted the “nominal” identification of the human good with *eudaimonia*, success or “happiness,” as also with *euzōia* and *eupraxia*, living successfully and acting successfully: *NE* I.4 1095a19–20; I.8 1098b2–22). So, according to her, if we are to defend Aristotle, we must establish connections between happiness and excellence of functioning *a posteriori*. At the very least we must make out the connection one way, and argue that a life of excellent human functioning is, as things currently are, the best way to secure the life that is best for a human—that in short “an enlightened prudence presupposes or requires morality” (p. 356): that is, as things are, “moral living” is necessary, even if not sufficient,

---

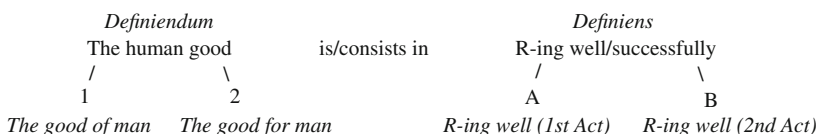
G. Lawrence (✉)  
University of California, Los Angeles, CA, USA  
e-mail: lawrence@humnet.ucla.edu

for living happily or as is good for a man, although, if things were to change, who knows?

These allegations are often run together, but mistakenly so. They differ over Aristotle’s *definiendum* and over his *definiens*, and so about the supposed gap between the two. Glassen’s claim is that, while Aristotle is aiming to define *the good of man*—identified with the final goal of man’s actions, the human end—he gets confused over the phrase “the good” (*to agathon*) and produces what is in fact a definition of *the goodness of man*, that is, of *what it is to be a good human* (a good one of our kind). Ackrill and Wilkes, by contrast, both suppose that Aristotle is aiming to define *what is good, or best, for a man*, the thing of greatest benefit to a human; but where Ackrill supposes Aristotle’s answer is ‘being the best possible man,’ Wilkes takes it to be “the life of the good man.”

These allegations are, I believe, variously the product of two main ‘ambiguities’ in Aristotle’s conclusion. The first is an ambiguity—or apparent ambiguity—in Aristotle’s *definiendum*, “the human good,” between a certain notion of the good *of* man and one of what is good *for*, or of benefit to, a man. If we take “the human good” as equivalent to the genitive phrase “man’s good,” then the ambiguity is between “what is the good of a man?” (i.e., what is his end, point, purpose, use) and “what is of benefit to a man, good for him?” (the difference being that between *possessive* or *subjective*, and *objective*, genitive constructions). The second is an ambiguity in Aristotle’s *definiens*, “<reason-involving> activity of soul in accord with excellence”—for short, “R-ing well, or successfully”—between so-called *first* and *second* actuality, that is, between the disposition to R well, and actively R-ing well (as for example in “Gavin sees well” between “Gavin is disposed to see well” and “Gavin is actually seeing well”).

It is simplest laid out diagrammatically.



As I view their interpretations, Glassen takes the *definiendum* to be 1 (the good of man), the *definiens* to be A (1st act); and the conclusion so viewed fallaciously to assert:

(i) 1 = A (i.e., *the good of man* consists in *being disposed to R well*)<sup>3</sup>

—fallacious because what A, as a first actuality, *properly defines*, or constitutes, is *goodness* in a human being, i.e.,

BGH (being a good human) = A.

This, while in itself correct, defines something other than the intended target, 1, and so, if viewed as Aristotle’s actual conclusion, would be irrelevant. Thus Glassen.

By contrast, Ackrill and Wilkes both take the *definiendum* to be 2 (what's good for man), but Ackrill takes the *definiens* to be A, and the conclusion to assert

(ii)  $2 = A$  (i.e., *what's good for a man* consists in *being disposed to R well*)

(Ackrill agrees with Glassen that  $BHG = A$ , that *being disposed to R well, or successfully*, is what constitutes *being a good human*, or goodness in a human; it is just that Ackrill takes this whole point as part of the right hand side, of the *definiens*, viz. that  $2 = (BGH = A)$ , while for Glassen  $BGH = A$  itself appears as the legitimate, but argumentatively irrelevant, conclusion.) In contrast to both Glassen and Ackrill, Wilkes, on the whole, takes the *definiens* as B—actually living the life of the good man—and the conclusion as

(iii)  $2 = B$  (i.e., *what's good for a man* consists in *actually R-ing well*).

For Ackrill and Wilkes the fallacy in the argument is that it is not obvious, respectively, what A, or what B, has to do with 2.

Of (i)–(iii), (iii) is the most plausible, both as regards truth and interpretation, and is, albeit with some charitable adjustment, what such defenders of Aristotle's argument as Kenny, Wilkes, and Whiting attempt, in their different ways, to make palatable.<sup>4</sup> But, in my view, none of the claims, (i)–(iii), actually follow from Aristotle's argument. However since I shall suppose here that Aristotle's conclusion asserts

(iv)  $1 = B$  (i.e., *the good of man* consists in *actually R-ing well*).

I don't see that concession as an immediate problem for Aristotle.

One further clarification is in order. Urmson, for instance, portrays the alleged fallacy as one in which the consideration of human function—of the “perfect specimen of a human life” (Urmson, p. 20)—will at best show that actual reason-involving activity, R-ing well, “is the good of man, not the good for man” (ibid). And objectors may indeed grant that it is plausibly true that  $1 = B$ , but claim that, since the *definiendum*, the good that is being sought (*to zētoumenon agathon*), is 2, the conclusion must, on pain of irrelevance, be supposed in effect:

$2 = (1 = B)$ .

We have here, they may say, another way of characterizing the fallacy, as in effect claiming that *what is good for a man is the good of man*, and an important one since it may explain how Aristotle came to commit it—being himself misled by the ambiguity in “the human good” between 1 and 2, and mistakenly taking the consideration of human function to clarify what is good for a man, i.e., 2, whereas all it clarifies is the good of man, 1.

By contrast I claim here that the *definiendum* is 1, as Glassen says, and that Aristotle's conclusion asserts  $1 = B$ . Of course I now have to make this plausible as an interpretation of the Function Argument. This is all I shall attempt here. But a fuller defense will require me to show that this interpretation doesn't save the

Argument at the price of making it irrelevant to Aristotle's project in *NE I* as a whole. In short, this paper is but Part 1, a *preliminary* foray, to begin sorting out certain muddles. Much of our difficulty in this area stems, I believe, from our lack of an accurate grasp on the grammar—the logical shape and category—of the locutions in the area, especially “the X-an good,” “good for,” and “the good of,” this in a teleological sense that seems largely to have fallen out of modern understanding.

## 20.2 The Basic Argument and the Inference Problem: Simple Inference or Principle?

In broad outline, the Function Argument's strategy is—and here I think Glassen is right (pp. 319–20)—(1) first to delineate the function of the human (1097b33–98a7), (2) then that of the good human (1098a7–15), and finally (3) to draw a conclusion about the human good (1098a16–18). And just as at the very beginning of the *NE* Aristotle had exploited a parallel between arts and chosen action, so here again he looks, if not wholly, at least initially and especially, to the *artisan* to parallel the human being. The argument that Aristotle seems to have in mind is this, taking the case of the Lyre-player:

- (L1) The function of the Lyre-player is playing the lyre.
- (L2) The function of the good Lyre-player is playing the lyre well or successfully, that is, in accord with the excellence(s) pertaining, or proper, to lyre-playing.<sup>5</sup>
- (L3) So the Lyre-player's good consists in playing the lyre in accord with the excellence(s) proper to lyre-playing.

Applied *pari passu* to the human case, this allows Aristotle to argue:

- (H1) The function of the Human is a specific way of being alive or living, viz. a kind of practical life of the part having reason (“R-ing” for short).
- (H2) The function of the good Human is R-ing well or successfully, that is, in accord with the excellence(s) proper to R-ing.
- (H3) So the Human good consists in R-ing in accord with the excellence(s) proper to it.

The schema of the argument is then, for functional item X:

- (X1) The function of the X is to *x*.
- (X2) The function of the good X is to *x* well or successfully, that is, in accord with the excellence(s) proper to *x-ing*.
- (X3) So the X-an good consists in *x-ing* in accord with the excellence(s) proper to *x-ing*.

There is some doubt, as we shall see, over the precise extension of X. But a necessary condition is being an item which has “some function and action” (1097b26).



What justifies the first move, from X1 to X2, and what is its point? Aristotle licenses it by appeal to a principle:

**P1:** the function of a good X is the same as the function of an X, but with the extra phrase “in accord with its excellence” added on to the function (1098a7–12).

—a principle he illustrates by appeal to the skilled (1098a9–12). The point of the move is less easy. It introduces *explicitly* the notions of “a good X,” of “*x-ing* successfully,” and the *equivalence* of this with “*x-ing* in accord with its proper excellence” (cf. a14–15). As a preliminary interpretation we can suggest that the thrust of this is (A) to bring out, or cash out, *the way* in which the X-an good and success of an X “rests in its function” (1097b25–27: see **P2** below)—that this is a matter not simply of doing what is for an X to do, viz. *x-ing*, but of course *x-ing successfully*, that is, *x-ing* as a good X, one with the excellence of an X, would *x*: so the functioning, or *x-ing*, at issue is *x-ing successfully* and this is equivalent to *x-ing excellently*.<sup>6</sup> (B) This thrust has an equally important *reverse* side. It had been a standard position to suppose that the X-an good, or success as an X, consisted simply in *being a good X*, in having the excellences of an X (cf. I.5 1095b29–96a2; I.8 1098b30–99a7; *EE* I.4 1215a20–25; Lawrence 2001, pp. 455–6). This view is now put in its place—the truth and the falsity in it separated out. The truth is that being a good human, one with the excellences, is *necessary* for attaining the human good, actually R-ing excellently, but it is not sufficient—because, for example, you can be a good human but asleep and not R-ing at all. The human good is a matter of *second actuality*, of realizing human excellence in actually living a human life excellently. So the move is *implicitly* clarifying the logical place of an older, “ancient” view about human success by refining our understanding of it via the distinction between first and second actuality (the critique implicit here is later drawn out explicitly at I.8 1098b30–99a7). We see both *that* the excellence(s) of an X, and being a good X (i.e., one with the excellence(s) of an X) do indeed have a role in an account of the X-an good, and *what* that role is, that is, how “logically” excellences come in to this account, as constituting the criteria of success in a thing’s functioning—cashing out what counts as doing it well or successfully: and it is such excellent functioning that constitutes the X-an good.

The move is *formal* to this extent: from the fact that it is the function of an X to *x*, it follows only that the function of the good X is to *x* well or successfully, where “to *x* well or successfully” is to *x* in accord with the excellences proper to *x-ing* (and Xs), *whatever these are*. It is no part of this move to determine what these excellences actually are, and so no part of the argument to yield a *substantial* determination of the good X, i.e., of the X with the excellences of an X, nor of what it is to *x* excellently. Of course it invites the question of a specification of these excellences as the natural next move; for this will yield a correlatively more substantial specification of the X-an good.<sup>7</sup> And in I.13 1102a5–8 Aristotle makes exactly this point, which then sets the agenda of the books that follow (*NE* II–VI).

Our puzzle, however, concerns the move from X1 and X2 to X3. What has our functioning as humans—or rather our functioning well as humans, our humaning, or R-ing, successfully, that is, excellently—got to do with our greatest good? This is a

“surprising step” (Kenny 1965–6, p. 54). In fact it may seem a blatant *non sequitur*, as it does to Glassen, Ackrill, and Wilkes, the last of whom remarks

...it is far from clear...how...the superb functioning of any *ergon*-bearing creature is relevant to what that creature’s greatest good is. (Wilkes, p. 343)

Why should what the good X does—viz. *x* excellently—determine, indeed constitute, what the X-an good is? Aristotle has already said that the human good, the X-an good where X is human, is generally agreed to be *eudaimonia* (I.4 1095a17–20), and he has argued that this agreement is correct (I.7 1097a25–b21). Now suppose, as Aristotle later claims, that the human excellences are the traditional virtues, and so the function of the good human would be constituted by acting *justly, courageously*, etc. Yet is such virtuous activity so obviously *necessary* in order to live successfully or happily? Many wicked people apparently flourish, and evade, in this life at least, any obvious comeuppance. A certain amount of circumspect dishonesty seems a reliable, though not infallible, way of securing a rather enjoyable, richer life, in possession of many of the truly good things of life—and if not the best life, at least one apparently more enjoyable than the impoverished and constricted life of Honest John on the dole. It may be that virtuous activity is necessary—that such dishonest people are not living happily even though they think they are. But it hardly seems *obviously* so. Conversely, is acting virtuously *sufficient* for living successfully, or happily? Surely acting justly or courageously will, in certain circumstances, require the sacrifice of life or the endurance of great torture, deprivation, and general misery, let alone minor annoyance and discomfort. Surely only someone set the task of “defending a thesis” in rhetorical debate would claim that *this* is their “greatest good,” or happiness.<sup>8</sup>

Admittedly, as already said, it is no part of the Function Argument itself to argue for *this* particular specification of the excellences. Yet this is the specification Aristotle will accept. Moreover won’t there be analogous worries for any specification of the excellences? For the gap appears to open between the formal notions of *living successfully* (or happily), and *living, or functioning, excellently*, because each can appear to be sensitive to *different* criteria, and so to be at best contingently co-extensional (cf. Wilkes’ claim).

However rather than cry *non sequitur*, let us ask instead what principle would license the inference to X3. A principle that would do the job is that:

Where the X is something with a function, the X-an good—the good or success, of the X as such—is, or consists in, X’s function—(i.e., in the X’s *x-ing* successfully or excellently).

However, once formulated, this looks like the very principle that Aristotle himself *explicitly* offers earlier on in the argument, where he says (1097b26–8):

**P2:** In the case of things that have some definite function and action, the good and the well/success (*to eu*) seem to be in the function.<sup>9</sup>

Now importantly, Aristotle’s point in saying this is *not*, as Glassen claims (pp. 319, 321), to give us “the drift” of the argument to come, as it were a preview

of it. Rather it is to give the argument's *rationale*—to *explain* why looking to human function might clarify the human good that is being sought (“*gar*” 1097b25). It is, admittedly, somewhat telegraphically expressed. There are two points of apparent unclarity. (a) First, **P2** says simply that “the good and well” of an X are held to “be in the function.” This leaves it unspecified, or only implicit, how exactly they are a matter of X’s function. As suggested above, it is this that **P1** and the move from X1 to X2 help to cash out. (b) Second, as we shall see below, there turns out to be a dispute over “the good and the well/success.” Some understand this as a claim about *goodness* in an X, or success as an X in the sense of *being a good X*, an X in a good state, and that it is *this* that is being said to be a matter of X’s function: that, for example, goodness in a sculptor or a pen is a matter of the sculptor’s sculpting well, and the pen’s writing well. That this, taken a certain way, seems a truism is taken as a welcome result, given that **P2** is presented as in itself obvious. However, as I interpret **P2**, it expresses the “Function-Good” principle that:

**FG:** Where the X is something with a function, the X-an good (X’s success) is in its function, in the sense of being a matter of its function, of its *x-ing*.

And in fact, as X2 makes clear, it is a matter of its *x-ing successfully* or well, and of its *actually x-ing* (second actuality). Success for an X is not constituted simply by being a good X (although that is a necessary condition).

So interpreted, **P2**’s role is precisely to license the inference from the specification of something’s function, X1, modulated by X2, to the specification of the good in its case, X3. (Joachim 1951, p. 49, and Reeve 1992, p. 128, also take this view of the role of **P2** in Aristotle’s argument.) So in the case of a sculptor, for instance, the sculptor-al good—success for a sculptor (*to eu*)—lies in the sculpting,<sup>10</sup> and consists in the sculptor’s actually sculpting successfully, i.e., in accord with the excellence(s) pertaining to sculpting. In his so acting, the sculptor is realizing and achieving his end, his good or success, *qua* sculptor (*viz.* fine sculpting).<sup>11</sup> So similarly if we can determine the human function, then a human’s doing that and doing it successfully, i.e., excellently—as the good human would do it—will constitute the human good and human success.

Yet if this is right, the objector may protest that not much is gained. The **FG** principle seems introduced more or less ad hoc to endow a fallacious inference with the aura of legitimacy, and the doubts that attached to the inference transfer simply to the inference principle. At best, the principle—introduced to give the rationale for the argument—so far from being self-evident, stands just as much in need of its own rationale. Aristotelian apologists will have to look elsewhere to defend it. Objectors, by contrast, will suggest Aristotle has misled himself due to an equivocation on “the X-an good” either in the principle itself and its use in the argument, or else in the relevance he takes the argument’s conclusion to have to illuminating the good he is inquiring after (*to zētoumenon agathon*). But, either way, we need to consider the **FG** principle more carefully, and also the nature of the good whose elucidation Aristotle suggests the Function Argument supplies.<sup>12</sup>

## 20.3 The Problem of the *Definiens*. Glassen's Interpretation of P2: The Good of Man Versus the Goodness of Man

It can, however, be objected that we are not yet at this point, on the grounds that **P2**—the claim that

**P2:** In the case of things that have some definite function and action, the good and the well/success seem to be in the function

—when properly interpreted, does not express the **FG**, function-good, principle (cf. p. 341 above).

Thus Glassen takes the move from X2 to X3 to show that Aristotle is confusing the notion of *the goodness of X* with that of *the good of X*. According to Glassen,

from the premise that **[H2]** the function of a good man is <rational> activity of soul in accordance with excellence, what follows is, not that **[H3]** the *good* of man consists in <rational> activity of soul in accordance with excellence, but that **[H4]** the *goodness* of man consists in <rational> activity of soul in accordance with excellence (p. 321, my H numbering).

That is, the conclusion Aristotle actually draws, **H3**, is illegitimate; the conclusion that the premises of the schema really license is **H4**. To explain this confusion, Glassen suggests locating it in **P2**. He argues that “the good” in **P2**'s “the good and the well/success” does not stand for the good of man—which “Aristotle has told us, is the final goal of man's actions” (p. 320; cf. 319)—but is the substantival use of the adjective “*agathos*” as that is used to qualify “lyre-player,” etc., i.e., the *goodness* of a lyre-player (the Penguin indeed translates it this way). If so, **P2** is not asserting **FG** (“a highly dubious proposition,” p. 322), but making a “truistic” point linking the function of things with their goodness, with their being a good one of their kind—viz. the **Function-Goodness** principle:

**FGness:** For all things, X, with a function, their goodness, their being a good X, “is in” their function, *x-ing*: that is, to be a good X ‘has to do with,’ or consists in, *x-ing* well or successfully, i.e., in accordance with its proper excellence (p. 322).

This principle licenses what Glassen takes to be the legitimate conclusion,

**[H4]** the *goodness* of man consists in <rational> activity of soul in accordance with excellence.<sup>13</sup>

Glassen's suggestion is that “the ambiguity of this term [*to agathon*] then misled Aristotle himself into treating it as if it referred to the final end of action,” and so fallaciously to draw **H3** as a conclusion, whereas all the argument licenses is the—irrelevant—**H4** (p. 322; cf. p. 321). That is, Aristotle is aiming to define the good of man, or the final end of human action, but all that really follows from his premises is a claim about human *goodness*, about what it is to be a good human.

However a closer look shows it is not Aristotle, but Glassen, who is confused. To take **P2** as Glassen does is to misunderstand Aristotle's *point* in asserting it. The context makes clear that **P2** gives the rationale for investigating function in order to clarify the good (or *to ariston*) in the "final end" sense that has been at issue. **P2** interpreted as stating the **FGness** principle clearly does not do this: it elucidates a different connection.<sup>14</sup> But there is, to my mind, a *definitive* objection to Glassen's interpretation.

Glassen evidently supposes that there is a legitimate inference from:

[H2] the function of a good man is <rational> activity of soul in accordance with its excellence,  
to

[H4] the *goodness* of man consists in <rational> activity of soul in accordance with its excellence.

via the inference principle, **P2**, understood as expressing:

**FGness**: the goodness of an X consists in—is a matter of—its functioning well.

The inference and its principle may seem, and are, unexceptionable—as witness:

[Pen2] the function of a good pen is to write well/writing well;

[Pen4] so the goodness of a pen, or its being a good one, consists in its writing well—is for it to write well.

This seems truistic—and Glassen welcomes that (p. 322; cf. p. 341 above). But in fact it turns on an equivocation.

Glassen overlooks the key point that, in Greek as in English,<sup>15</sup> there is an ambiguity in the phrases "to  $\phi$ " or " $\phi$ -ing" between first and second actuality. That is, "Gavin sees" or "plays chess" may be used to say that I possess the abilities to see and to play chess (first actuality); or else that I am currently seeing or playing (second actuality), a sense where in English it is often natural to use the form "is  $\phi$ 'ing" (cf. p. 336 above).

Now the sense at issue in the **FGness** principle, if it is to be true, is that of *first* actuality. Thus when we say that the goodness of a pen consists in its *writing well*—or that a good pen *writes well* (or is *one that writes well*)—the sense properly at issue is that of first actuality: that is, a good pen is one that *can*, or is disposed to, write well—is one that has the excellent states proper to writing.<sup>16</sup> After all, a pen that isn't being used but is in a drawer doesn't thereby lose its goodness or cease to be a good one. Compare Foot's remark in "Goodness and Choice":

Where a thing has a function, the primary. . . criterion for the goodness of that thing will be that it *fulfills its function well*. Thus the primary criterion of goodness in a knife is *its ability* to cut well. (1961/1978, p. 135)

As the illustrative gloss shows, Foot uses the phrase “it fulfills its function well” in the first actuality sense of “is able to fulfill its function well,” and not the second actuality sense.<sup>17</sup> The goodness of an X then consists not in its actually *x-ing* successfully, or excellently, but in its being so disposed. If so, the sense of “rational activity done successfully/excellently” that is required for the truth of **(H4)**—the claim that

[H4] the *goodness* of man consists in <rational> activity of soul in accordance with excellence

—is that of first actuality, of being disposed to R excellently (i.e., having the pertinent excellence(s)).

But—and here is the nub—the sense of “rational activity of soul in accord with excellence” at issue in the specification of human function in premise **[H2]**—and so in Aristotle’s conclusion **[H3]**—is, as Aristotle *explicitly goes out of his way to clarify*, that of *second* actuality (1098a5–7).<sup>18</sup> Yet if this is the relevant sense, then to draw **[H4]** from **[H2]** would involve a category mistake: for the goodness of something consists in a disposition, i.e., a first actuality, not a second. Indeed **[H4]** only *seems* to follow because Glassen is misinterpreting **[H2]**, equivocating on the relevant phrase, incorrectly taking it in the sense of “consists in *being able* to engage in rational activity well.” It is Glassen here who is guilty of trading on an ambiguity.<sup>19</sup>

There is indeed something rather odd about Glassen’s mistake. For his cry of fallacy to work, he must be assuming that the good of man cannot consist in the goodness of a human: otherwise “R-ing well” in correctly constituting human goodness could equally constitute the good of man (or the final end of action). The reason this assumption is correct turns on the first/second actuality distinction—the two senses of R-ing well. But if you could sense that, you are equipped to appreciate the right way of taking the argument and so avoid this allegation of fallacy!

*To sum up.* If I am right,

- (a) It is not Aristotle who is confusing the notions of the good of X and the goodness of X; and we have been given no reason to suppose he uses “the X-an good” (*to X-an agathon*) in what Glassen calls the substantial sense of “the goodness of X” here or indeed ever (*aretê* plays that role).
- (b) “The good and the well/success” in **P2** is to be interpreted, not as the goodness of X, but as *the X-an good and X-an success*, and this removes a block to construing **P2** as **FG**. The phrase is still open to our original ambiguity, between “good of” and “good for,” but a potential further, third, ambiguity, “the goodness of,” can be dismissed at least in the present context. If so, Aristotle is not making a careless slip in moving from **H2** to **H3**. He thinks, rightly or wrongly, that this is licensed by an acceptable principle. (Whether it is so, we need to consider.)
- (c) Finally, this clearly resolves our question of the *definiens* in favor of **B** over **A**. Interpreters who treat Aristotle as offering an answer in terms of first actuality—of being a good human, or in being disposed to R well—simply go against the

text, and in effect adopt that “wise and rather ancient” position which Aristotle himself is at pains to criticize in *EE* I.4 1215a20–25 (cf. *NE* I.8 1098b30–99a7; Lawrence 2001, pp. 455–6).

## 20.4 The Problem of the *Definiendum*: Is The Human Good What is Good for a Man?

So far so good. The *definiens* is clarified and we have banished one misconception of “the good.” Yet the notion of “the human good” (*to anthrôpinon agathon*) which is Aristotle's target, and, generally, “the good” at issue in **FG**, which I take to be the generalized version, i.e., “the X-an good,” are still not immediately transparent notions. (1) What, for instance, is the extension of X in “the X-an good”? (2) And what is this sense of “the good”? Such substantival uses of “good” as “the good” have attracted little modern attention, or rather been avoided. Geach (1956/67) is concerned only with adjectival uses; Ziff (1960) mentions the use, claiming that

Etymologically speaking, such morphological constructions as ‘The Good,’ “goodness,” “goodity,” are derivative from predicative or attributive occurrences of “good”....

and then proposes to ignore such constructions (pp. 209–210). Vendler (1963) too focuses only on adjectival constructions, as does Thompson (1994). (Von Wright 1963 does address such locutions. But his account turns out to be somewhat idiosyncratic.)

How then are we to understand “the human good” that is the target of Aristotle's elucidation? Well, we know what Aristotle takes as an initial—if somewhat “nominal”—*answer* to the question he is posing: almost everyone would agree, he says, that “by name” *eudaimonia* is the human good, and *eudaimonia* is taken as synonymous with *eupraxia*—doing or faring well or successfully—and *euzôia*—living well or successfully. So to ask after the human good is equally to ask after *the good, or the best possible, human life*. It is the answer to this that Aristotle takes to be controversial (I.4 1095a20–28). But what actually is the question? One way of answering it according to Parfit is offered by an *Objective List Theory* which specifies certain things as *objectively* “good, or bad, for people,” i.e., irrespective of their actual, subjective, desires about them (1984, pp. 493, 499). And indeed many interpreters of Aristotle—detractors and defenders alike (e.g., Ackrill, Kenny, Wilkes)—suppose that, in asking after the human (or X-an) good, Aristotle is asking, and asking objectively:

“What is *good for* a human being (or X)?”

And it is natural to take this, as Whiting does, as equivalent to the question

“What is beneficial for, or of benefit to, a human being?—what does them good?”

(Cf. Ackrill 1973, p. 20; Wilkes 1978, p. 356). Thus Whiting invites us to consider the ambiguity Aristotle finds in the notion of *hou heneka*, as this appears in the sentence, “x is for the sake of y”:



One is *the beneficial sense* in which x's occurring benefits someone. The other is *the instrumental sense* in which x is instrumental (or a means) to bringing it about that y, and it is a further question whether anyone is benefited by the process: it is simply a matter of causal efficacy. (p. 35, my italics.)<sup>20</sup>

Her idea is apparently that we should gloss the notion of the human, or X-an, good in the first of these senses, as “that which is for the sake of a human, or X, *as beneficiary*” (pp. 37, 38). For this is clearly the more plausible of *these* two senses in this context. So when Aristotle asks what the human good is he is asking what is most of benefit to a human. If so, then when Aristotle gives the nominal answer as *eudaimonia* (or *euzōia* or *eupraxia*), we are to understand this as claiming that the greatest good *for* a human, or what *benefits* a human most, is a wonderful life—to live successfully, or well.

If the human, or X-an, good is to be understood this way, as *the beneficial good*, and if **P2** is to be taken as expressing the function-good inference principle, **FG**, then **FG** must be construed as:

[**BenFG**] Where X is something with a function, the X-an good, in the sense of what is good for, or benefits, an X, consists in X's doing its function successfully or well.

*However* this interpretation of the human, or X-an, good, and of the **FG** principle, is problematic. For a start, as Wilkes says,

... it is far from clear... how... the superb functioning of any *ergon*-bearing creature is relevant to what that creature's greatest good is. ... (Wilkes, p. 343)

There seems to be a gap between the life of the good man, i.e., a life of excellent human functioning, and the life that is good for man. Aristotle appears to be sliding between two senses of “doing well” or “living well”:

... a man may do or live well, in the sense that he performs admirably the activities that his *ergon* ascribes to him, with or without doing well for himself or living a life that is good *for* him. (Wilkes, p. 343, her emphasis.)

Of course, if **P2** does express **BenFG**, then Aristotle explicitly asserts a principle of connection that, formally speaking, closes this gap and removes a possibly fallacious inference. But then why accept the principle? If **P2** is supposed to provide the rationale for the Function Argument, it is troubling if this rationale is not self-evident but itself in turn stands just as much in need of its own rationale. Moreover the problem is not simply lack of self-evidence. For if X in “the X-an good” is a beneficiary, then various counterexamples apparently make it clear that, to be plausible, the range of possible substitutions for X in the X1–X3 argument-schema must be a smaller class than simply functional items, and the principle **BenFG** must likewise be restricted in its scope.

Let us turn to these counterexamples.

(a) *The Alleged Counterexamples*

We are offered a range of counterexamples. Wilkes offers the sheepdog:

**Example 1:** What the good sheepdog does and what is good *for* the sheepdog to do have no necessary correlation. (p. 346)

Kenny offers the sculptor:

**Example 2:** Surely, we feel inclined to object, what is good *for* sculptors (e.g., adequate remuneration and good living conditions) is quite different from what the good sculptor does (e.g., sculpt well). (p. 27)

Whiting (1988, p. 33) gives her version of the two moves of the argument's schema from X1 to X2 and X2 to X3,<sup>21</sup> and then remarks:

The legitimacy of these moves is typically challenged by appeal to the following sort of example. [A] From an understanding of the function of a knife, it may follow that being sharp and cutting well make something a good knife: but it does not follow that being sharp and cutting well is *good for* a knife. [B] Similarly, from an understanding of what it is to be a flute player, it *may* follow that some things (e.g., perfect pitch and a sense of rhythm) make someone a good flute player: but it does not follow that these things are *good for* someone who plays the flute. In a depressed economy, an unemployed virtuoso may wish that he had been tone deaf and instead become a doctor. (p. 34)

This is difficult to disentangle. The range and type of subject that is at issue for being "what is good for an X" is confined (a) to the functioning of the X and (b) that taken as actual functioning, i.e., second actuality ("their performances;" cf. Ackrill quoted in n. 23). Let us simplify and emend Whiting's statement so that her challenge runs:

**Example 3:** It does not follow from the fact that the function of a (good) knife is actually cutting (well) that cutting successfully or well is good for, i.e., benefits, a knife; nor

**Example 4:** that actually playing the flute successfully or well is good for, i.e., benefits a flute player.<sup>22</sup>

Thus, taking *x-ing* to be the function of the X, the schematic form of the worry, at least when cleaned up, is thus:

(W) It is not obvious why *x-ing* well (second actuality)—successful or excellent functional performance—is what is good for, or of benefit to, X.

The function of the X, *x-ing*, is *the exercise of X's essence, or quasi-essence* (in the case of artisans and artifacts): an X's actually doing what it is for an X to do. The worry is that there seems no obvious reason why an X's doing this (*x-ing*), and doing it well, should be of *benefit* to an X—and given the counterexamples reason to think that in many cases it isn't, or makes no sense.

The counterexamples offered are of several sorts: people in technical roles; animals in technical roles; artifacts. To these we can add the case of parts of the body. What sense is there to the claim that seeing well is good for, or benefits, the eye?<sup>23</sup>

Now both Whiting and Kenny try to defend Aristotle against this difficulty.

(b) *Whiting's Response*

Whiting's strategy is to concede the force of these counterexamples, but then to sideline them by arguing that Aristotle disallows precisely such cases as legitimate substitution instances of the X1–X3 schema of inferences (or rather her version of that). Legitimate substitution, she argues, is restricted to the case of animate natural kinds, and thus excludes both artifacts ("inanimate objects") and artisans ("non-natural kinds"). She also offers an explanation of why Aristotle should accept the validity of the inference in the favored cases.

Her argument against the counterexamples (pp. 35–6), as I understand it, is this:

- (1) Aristotle distinguishes two senses of "*hou heneka*," of "x being for the sake of y," *the beneficial* and *the instrumental*. It is the human good in the sense of "what benefits a human" that is Aristotle's concern here (cf. pp. 37–38).
- (2) Aristotle restricts the class of beneficiaries to living creatures (a claim she bases on *NE VIII.2 1155b29–31*).<sup>24</sup>
- (3) So this rules out inferences from the X2–X3 schema where X is an artifact, like a knife.
- (4) To rule out the cases of artisanal ("nonnatural" living) kinds, like the flute player and the sheepdog, Aristotle must show further that "the move from <X2 to X3> is warranted only in substitutions in *natural* <living> kinds." And to show this is not *ad hoc*,

Aristotle must establish some connection between a thing's membership in a natural kind and what is beneficially good for that thing—or... some connection between a thing's essential properties and what is beneficial for that thing. (p. 36)

And Whiting then goes on to characterize this connection in terms of a distinction between conditional benefits and categorical benefits.<sup>25</sup> Whether or not playing the flute is good for, or benefits, me is conditional on the (subjective) desires and interests I happen to have (e.g., if it is "my sole source of support or personal fulfillment"): whether playing the flute benefits me, or is in my interest, depends on whether it is an interest I happen—subjectively—to have. By contrast, for a (normal) member of a species to realize its end—in living the life that is for it to live—is *categorically or unconditionally good for it*, "that is good for it whatever its actual interests and desires" (p. 36), i.e., objectively. This way of ruling out the cases of nonnatural living kinds involves Whiting in articulating the validity of the principle of inference for living natural kinds, and then defending it in terms of Aristotelian commitments to essentialism and final cause explanation (cf. Reeve 1992, p. 128).

Thus the way she sidelines the cases of artifacts and of artisans (her "nonnatural kinds") are very different.

Her way of trying to deal with the counterexamples is, I believe, misguided. For a start, too much is asked of the reader. In the argument itself Aristotle makes no move to restrict the proper substitution instances only to functional items of a natural kind; nor to explain the theoretic basis of such a principle. On the contrary, he feels free to appeal precisely to the supposedly controversial examples—artisanal kinds and body parts in the *NE*, artifacts in the *EE*—to illustrate the argument without any explicit suggestion that they are illustrative only of limited points in the argument.<sup>26</sup> *Above all*, the very principle of the inference, **FG**, that links the function of X with the X-an good, is introduced *explicitly* by way of examples of nonnatural kinds (artisans)—and possibly also artifacts:

Perhaps then this [*eudaimonia*/the human good] would become <clearer> if the function of the human were grasped. For just as for a flute-player and a sculptor and for every craftsman, and *generally for things [hôn] of which there is some specific function and action*, the good and the well/success are held to be in the function, so it would be held also for man, if there is some specific function of him. (1097b25–28)

If “*hôn*” is neuter, as I believe it is, then the principle is being generalized over all functional items, including artifacts; but even if it is masculine, and is generalizing only over humans with some sort of role, wider than the “technical,” yet one can hardly suppose Aristotle is cutting out artisans as legitimate substitution instances in this principle!<sup>27</sup>

Now, admittedly Whiting doesn't agree that **P2** expresses **FG**. Like Glassen, she apparently supposes it concerns *goodness* in an X (see n. 21). This means that the principle underwriting the inference from (good) human function to human good, must, on her view, only be *implicit* in the argument. This gives her the freedom to suppose that Aristotle adheres to the **BenFG** principle only in the restricted form:

[**NatBenFG**] Where, and only where, X is a living natural kind with a function, the X-an good, in the sense of what is good for, or benefits, an X, consists in X's doing its function well or successfully.

Of course this itself needs defense, as Whiting acknowledges and tries to provide.

But again this is, I believe, too much to expect of the reader—who now has to supply both the needed principle of inference, and the defense and explanation it needs. Moreover, as I have argued in Section 20.3, I think that *given its context*, **P2** *must* be taken to express **FG**, not **FGness**. If so, then given **FG** is explicitly supposed to hold of artisans, we simply *cannot* sideline them in the way Whiting proposes, and must find another way out of the absurdities in which it appears to land us.

In fact, I think, the proper treatment of the alleged counterexamples demands a double response. For, as I will argue, there are *two* mistakes that get made here.

### (c) *Kenny's Response*

The first is revealed by Kenny. He points to a different way out. He acknowledges our worries about the inference from X2 to X3:

Surely, we feel inclined to object, what is good for sculptors (e.g., adequate remuneration and good living conditions) is quite different from what the good sculptor does (e.g., sculpt well). (1965–6, p. 27)

but then responds:

But presumably Aristotle would reply that this merely showed the difference between what was good for sculptors *qua* men and what was good for them *qua* sculptors. What is good for a man, *qua* man, to do is what the good man in fact does *qua* good man. But what the good man does is what all men should do. . . .

This point—“the *qua* point”—seems completely correct. A thing may have several different functions, and it is important to keep the inference schema relativized to the function in hand and not equivocate on it, as happens in some of the alleged counterexamples, such as Wilkes’ sheepdog and Whiting’s flautist, and (intentionally) in Kenny’s own example of the sculptor. (Neither Wilkes nor Whiting seem fully to appreciate this).<sup>28</sup>

But, although correct, Kenny’s point does not go far enough. This is clear from the case of artifacts, which, like the knife, lack any such obvious equivocation. If we are worried, as we should be, about saying that its activity, *cutting well*, is *good for*, or of benefit to, a knife, it doesn’t seem any help to add “well, good for a knife *qua* knife.” (We are not concerned here with the possibility of an instrumental loop—where its cutting well happens to keep the knife sharp, and so ensures that it cuts well when used to cut). And *pari passu* with body parts.<sup>29</sup>

But the problem doesn’t stop with artifacts and body parts. For in the examples of the sculptor, sheepdog, and flautist, even when we correct them and take the issue to be not what is good for the sculptor *qua* man, et al., but what is good for, or of benefit to, them *qua* sculptor, et al., the question arises of whether there is any such thing. Indeed what does it even *mean* to talk of *what is good for* the sculptor as such, or the sheepdog as such? The obvious sense is *what would (instrumentally) help* the sculptor attain their end, or activity, *qua* sculptor. But clearly *that* is not at issue. Kenny’s suggestion is apparently that what is good *for* the sculptor *qua* sculptor to do is what the good sculptor in fact does *qua* good sculptor—that is, good sculpting, their end or activity. But in what sense does good sculpting *benefit* a sculptor as such? It is not that thus they earn money or fame—for such things do not belong to the art of sculpting, *as such*. Good sculpting is success, or perfection, as a sculptor; it is their end as such, it is the good *of* a sculptor as such. But is attaining, or realizing, this end—their good functioning—good *for* them *as such*? It is not obvious what sense such a claim has (other than the possibility of an instrumental loop).

We seem here in the territory of the dispute in *Republic I* between Socrates and Thrasymachus, over who benefits, or gains advantage, (*sumpheron*), from the practice of a skill *as such*—the practitioner as such, or the object of the skill (*Rep.* 341c4–342e11, etc.). A skill, if it is perfect (*teleia*), and its artisan, if perfect, stand in no need of any benefit (*Rep.* I 342c4–6), any more than god—the object, or end, of our worship—is a *beneficiary* of it, being already a perfect being in need of nothing (*EE VIII.3* 1249b13–16; cf. *Euthyphro* 13a–15a). There is no way to benefit something that is already perfect; for *qua* perfect, it stands in need of nothing.

But doesn't this miss the point? Isn't the point that it is a thing's perfection that benefits the thing—indeed is its greatest benefit? Of course, once perfect there is, as such, no *further* way left for it to be benefited?

But surely good doctoring—success or perfect actualization as a doctor—benefits the patient as such, and not the doctor as such (though, incidentally he may be the patient (cf. *Phys* II.1 192b23ff). How would good doctoring—perfect actualization as a doctor—*benefit* the doctor as such, i.e., as a practitioner of the medical art as such—instrumental loops aside?<sup>30</sup> And whether or not this is universally the case, such examples are enough to throw into doubt the idea that generally their proper activity, their functioning successfully, is something that can be said to benefit, or be good *for*, things with a function.

But perhaps this is the very line of contrast between artisans and natural organisms. Nature faces inwards, while art faces outwards to the deficiencies of nature in others. The functioning well of organisms, their doing their proper/essential activities, is what is good for, or what benefits, *them*, whereas the functioning well of artisans, their doing their proper/essential activities, benefits *others*.

(a) Yet does this really hold up? Why not say rather: with natural organisms operating naturally, they just do their thing—doing that well or successfully doesn't benefit them, it just *is* their good, it doesn't *do* them good. But art, art steps in where nature fails—to *help* nature secure its own goals: it benefits things that have natures, where nature has fallen down and needs aid (crudely speaking). Art benefits, because art is parasitic and other-directed, but nature is “at home” and not in the business of promoting or benefiting anything, just of being itself and realizing itself. (b) But, *even* were the above idea to hold up, and to meet the challenge of *giving a sense* to an organism's functioning successfully being of benefit to it—to its being *the beneficiary* of its functioning—*nonetheless*, given that (1) **P2** explicitly applies also to artisans, and that (2) it expresses **FG**, then (3) the X-an good at issue here in **P2** *cannot* be interpreted as the beneficial good—simply on the grounds of this artisanal parallel with which it is explicitly introduced (unless, that is, a sense of “beneficial good” can be supplied equally for the artisanal case). And the point is even clearer if, as I believe, the **P2** principle applies also to artifacts, like knives.

This difficulty should make us query whether we have correctly understood the human, or X-an good, that is the target of elucidation in the Function Argument: is Aristotle's concern really with what is good for, or benefits, an X? At the same time, the above remarks point to a different way of conceiving the good in question—not as what is good *for* an X, but as the good *of* an X.

But before that, I want to dwell a little further on the notion of the beneficial good.

(d) *Further Queries about “Good for” and “the Beneficial Good”*

This notion of “what is good (or bad) for an X” is somewhat obscure. A natural, and relatively unproblematic, way to understand it is as equivalent to the beneficial,

and to understand the latter, along with its correlative, the *harmful* or *deleterious*, as broadly instrumental, or end relative, viz., very roughly:

B benefits/harms the X in that B, directly or indirectly, promotes/impedes the attainment of the X's proper ends, or proper functioning.

Von Wright, for instance, takes this instrumental understanding of the category of the beneficial/harmful (1963, pp. 41ff) (although he takes the category of the beneficial as a sub-form of the useful—"utilitarian goodness"—as being what is favorably causally relevant specifically to *the good of some being*, *ibid.*, p. 42). And Aristotle certainly appears to have such a category (cf. I.7 1097a26–7)—of the helpful, the *ôphelimon* (cf. *NE* I.6 1096b10–16), the expeditious, or advantageous/harmful, the *sumpheron/blaberon* (II.3 1104b30–1105a1), the useful, the *chrêsimon* (VIII.2 1155b18–21; see also Plato *H. Ma.* 296d–297d; *Rep.* II 357b4–d2; cf. Von Wright, *ibid.*, p. 41).

So viewed the X's end, or its good, or success, sets the *horos*—the criterion—for what counts as beneficial or harmful. It is in this vein that Aristotle claims, in *NE* VII.13 1153b17–25, that human *eudaimonia* (success/happiness) is the criterion for goods of the body, and goods that are external and a matter of chance. For these are needed because, but *only to the extent that*, their presence promotes and enhances, and their absence impedes, the activity (or activities) that constitute human *eudaimonia*. We can view this too as an **FG** principle, moving from a thing's function to its good, but in the very limited sense of what is instrumentally good for, or harmful to, it:

[**InstBenFG**] Where the X is something with a function, the X-an good, in the sense of what is *instrumentally* good for, or beneficial to, the X, is set by X's function.

Thus the function of eyes, seeing, determines the ocular good in this sense of what is good for the eyes, i.e., of what instrumentally aids their (successful) function, or impedes it.

But clearly it is not, or not simply, this basically instrumental sense of good for, or beneficial, that could be at issue in the Function Argument. Aristotle is *not* seeking here to clarify what is beneficial for humans in the sense of promoting, or being advantageous, to their end, but to clarify what that end is (1097b22–5). (Even Prichard (1935, in 1949, pp. 51, 53) who claimed that the former was really Aristotle's question had to admit that, if so, Aristotle here in the Function Argument expressed himself "in a misleading way" and "misrepresented his own view"! Prichard's thesis was demolished by Austin 1967.)

So, this still leaves us with the challenge of what sense can be given to the notion of what is good for, or beneficial to, an X in **BenFG**. What is this sense in which Ackrill and Wilkes find the claim that "X's *functioning well* is good for, or of benefit, to it" not self-evident, and in which Kenny and Whiting try to defend it?

Scholars often talk here, and elsewhere, in an undifferentiating way about "the good for Xs" and "what is good for Xs" (cf. Reeve 1992, p. 128; Urmson 1988,



p. 20). In these phrases the “for” is ambiguous between what in classical languages used to be called a dative of *advantage* and a dative of *respect*: that is, it is ambiguous between the notions of what is good for, or benefits, Xs and the good, with respect to, or *in the case of*, Xs (Hurka 1987, p. 73, senses something of this ambiguity). We consider this latter notion in the next section. So the dilemma I pose is this:

- (1) If we take the “for” as a dative of advantage, the sense of the question “what is good *for* a human?” appears to be instrumental—and that is not what Aristotle is asking;
- (2) If we take the “for” as a dative of respect, the sense of the question is “what is the good in the case of a human?”, and this is to take the human good as the good *of* a human, and not per se to raise questions of benefit and beneficiaries at all.<sup>31</sup>

One way the opponent can respond to this challenge is to claim that what is good for, or beneficial or advantageous, to an X encompasses more than what is merely instrumental to X's (proper) ends. Thus Brad Hooker, for instance, distinguishes between “instrumental” and “constitutive” benefits:

Something is instrumentally beneficial to someone if it is a means to some further thing that itself constitutes a benefit to that person. . . . Different theories of individual welfare—of what makes a person's life go well or badly for him or her—differ over what things *constitute* benefits to people. (1996, p. 141)

I presume Hooker's idea here is that something may be a benefit to an X not by being an instrumental means, but by constituting their life going well or badly. This seems at least what, nearer home, Whiting has in mind when she claims that Aristotle's view

is that for each species there is an ultimate end such that realizing that end (which Aristotle identifies with living a certain sort of life) is categorically or unconditionally *good for* any normal member of that species—that is, *good for it* whatever its actual interests and desires (p. 36),

or as she also says “*intrinsically and non-instrumentally* beneficial” for it (p. 37). According to Whiting, as we saw earlier (p. 11), such benefits can only be enjoyed by living natural kinds—only they can be beneficiaries. So Aristotle is committed to **NatBenFG**—or as Whiting expresses it, the claim that

something's membership in a natural kind at least partly determines what is beneficial for that thing (p. 38)<sup>32</sup>

and this in her sense of “categorically good for it,” and not (merely) **InstBenFG**. Thus, she says, a plant's activity of photosynthesizing is part of the plant's healthy functioning and as such good for it.<sup>33</sup>

But is this healthy functioning good *for*, or of *benefit to*, the plant? Here is this plant doing its planty thing—realizing its nature: is that good *for* it, of benefit to it? Certainly you could harm or damage the plant by impeding this activity. Yet it seems that if one says that (good) photosynthesizing is good *for* the plant, the natural way

to take this is that photosynthesizing achieves certain other results that speak to the plant's needs.

My stronger challenge then is that the idea that an organism's realizing its end, i.e., its functioning successfully, is good for, or beneficial to, it—"intrinsicly and non-instrumentally" beneficial to it—as yet lacks a sense; my suspicion is that none will be forthcoming (unless being of *benefit to* something can be heard as being to its good, and this as being part of or constitutive of its good: but this is an extension I find hard to hear in the natural profile of the concept: see further below). Can anything *benefit* from attaining its own proper end? We found it difficult to see what sense it made to say that good sculpting, or good sheep-dogging, was *good for* sculptors, or sheepdogs, as such. The same puzzle occurs with living natural kinds: what sense—instrumental loops aside—does it make to say that an organism's living the life-that-is-for-it-to-live is good for, or benefits, it? It is the good *of* it, but is it good *for* it? It *is* its good, its success, but does it *do* it good?

Initially one might think: "There's really no puzzle. What greater benefit can you give someone than a long happy life? Isn't this the best present the fairy godmother can give—the greatest benefit? And what greater harm can you do to someone than deprive them of this?"

But the good fairy helps you by arranging that the circumstances of your life be such as to allow you to attain (and enjoy) the end of being human (the good of you, in a sense to be explained), and by making sure that those circumstances which would impede or prevent you attaining your good do not arise. She helps to secure—to promote and protect—the human good in your case: and does so for your sake—you are the beneficiary of her actions, while the end of her actions is your *eudaimonia/ euzôia*. But this doesn't show that your living or functioning well (your *euzôia*)—your welfare or faring well—itself is good for you, or benefits, you. She is benefiting you by promoting, or securing, your good, not by making your good, your living well, good for, or of benefit to, you.

After all, if one said "her noble life was good for her," or "her contemplative life benefited her," it is hard to see what this could mean, other than that it promoted some other goods. Or again, consider an analog with the living body. Health is the functioning well of the body; and it determines—is the *horos*—of what is beneficial and harmful to the body. It is success in a body—its end, the good of the body: but what would it mean to say that its healthy functioning is good *for*, or benefits, the body as such—"intrinsicly and non-instrumentally good for it"?

The topic is difficult. But as yet I remain unconvinced that the notion of "what is intrinsicly and non-instrumentally beneficial to, or good for, X" has been given a sense. Its defenders presumably would not want to say this is just a baroque way of talking about the good *of* Xs, their final causes. And indeed they cannot. For in their intended sense of intrinsic benefit only living things can be beneficiaries, whereas, as we shall see, we can talk of the good *of* X in the case of *any and every* functional item.

Suppose, however, that I am wrong, and my challenge can be met—that there is, after all, an intelligible question, at least in certain cases, over whether the item's successful functioning is of "non-instrumental" benefit to it. (That Aristotle *does* so

think may be suggested by the remark at *NE* IX.8 1169a12–13.) My second, weaker, claim is that, even so, it is not the human (or X-an) good in this “beneficial” sense that can be at issue here in the Function Argument. If, as I believe, the X-an good at issue encompasses items, like the knife, for which this question of beneficial good is agreed not to arise, then it seems either Aristotle is guilty of adopting a fallacious principle or these objectors and defenders both err in their common interpretation of the X-an good. I opt for the latter.

To recapitulate. I think that the idea that X's successful functioning is, at least in certain cases, good for, or of benefit to, it *in some non-instrumental way* has not been given a sense. And even if it has a sense, it is not such a sense that is at issue in the Function Argument.

## 20.5 An Alternative Interpretation of the Human Good: The Good of Humans

There is another way to understand “the X-an good,” which avoids the above difficulties, and which allows us to understand FG in the unrestricted way in which it appears to be presented.

As I said above, if we take the notion of the human, or X-an, good as one of “the good for humans/Xs” this is ambiguous between a *dative of advantage* and a *dative of respect*. This latter notion—the notion of what is the good with respect to, or in the case of, Xs—is as yet obscure. Indeed perhaps one way to understand it is as asking after what is *good for*, or what benefits, an X. But there is another way, for which I shall reserve the locution “the good of X.” To ask after the X-an good in *this* sense is to ask after X's end or point (cf. “What's the good of it?”): this is thus a *teleological* notion of the good (cf. Lawrence 2006, pp. 39–41).<sup>34</sup>

To ask after the X-an good, or the X's good, would then be ambiguous between asking about what is good for, or benefits, Xs, and asking about what the good of Xs is, their end or point. For example, *the ocular good* could be either that which is good for, or beneficial to, eyes, or else what constitutes the point or end of eyes.

The X-an good, in this sense of the end or *telos* of an X, is X's final cause. It is “that for the sake of which [*hou heneka*] the rest [*ta loipa/ta alla*] are” (cf. I.7 1097a18; cf. *EE* I.8; *EE* II.1 1219a8, a10–11), where “the rest” are everything else that properly speaking have to do with X. Aristotle here employs that sense of “*hou heneka*” which he elucidates as “that for the sake of *which, as end*” (*hōs telos*; *EE* I.8; cf. *Phys* II.2 194a27; II.3 194b32ff) as against the other sense “that for the sake of *whom, i.e., to benefit whom*” (which he sometimes glosses with the dative of advantage “*hō(i)*”).<sup>35</sup> So, the X-an good in this sense is the final cause of X.

This interpretation is helpfully highlighted by its contrast with Whiting's. She takes the ambiguity Aristotle finds in *to hou heneka* to be between “the merely instrumental and the beneficial senses in which we say that one thing is for the sake of the other” (p. 35). To say this is to suppose, in effect, that the contrast is between two senses, or completions, of “*good for*”—between what is good for achieving

or promoting an end, and what is good for someone (cf. n. 21). But the ambiguity Aristotle is after lies in the “*hou*” of the *hou heneka*, between a masculine and neuter reading: between asking “for the sake of *whom*?”—i.e., “to whose benefit?”—and asking “for the sake of *what*?”—i.e., “to what end?” Aristotle’s distinction is thus between *end* and beneficiary—not *means* and beneficiary.<sup>36</sup> (Thus god, for example, is that for the sake of which we worship, as the end, not the beneficiary, of our worship: cf. *EE* VIII 1249b13–16.)

(i) *Final Cause as the Principle of Organization*

The X-an good in the sense of X’s end or point is that for the sake of which everything else that has to do with, or belongs to, X is. It is thereby equally the *principle of organization* for everything else that has to do with, or belongs as such, to X. Consider skills, and the skilled. Aristotle says (*NE* 1097a18–22),

What is the good [*t’agathon*] of each [action and *technê*, or of the just mentioned skills]? Is it not *that for the sake of which the rest* [*ta loipa*] is done? And this in medicine is health, in generalship victory, in house-building a house, different things in different things—and in every action and choice, the end: for it is for the sake of this that all do the rest [*ta loipa*].

The good of medicine—“the medical good”—is its end, health; and it is around this end that everything else in medicine is organized (has its rationale); if it lacks an appropriate connection with health, then it has no place in medicine. Or consider artifacts. Thus the good of a knife—the knife-an good—is its end, cutting. It is this then that is the organizing principle of everything to do with the knife. (When we ask “what’s the good of a knife?” we are asking what its point is, i.e., we are assuming *that* it has an organizing principle—and is not a mere accidental unity—and asking *what* that is). It is with reference to this end that various dispositions count as excellences or defects of the knife—so it is this that provides the criterion of what is to count as a good and bad knife. And it is with a view to this end that it has the parts it has, is made of the material it is, is properly held the way it is, needs to be kept in certain ways and not in others, etc. So, unsurprisingly, the good of X, or its end, is the measure for what is good and bad for it, beneficial and harmful (e.g., being used on stone or left out in the rain). That is, the principle that links function with instrumental benefit, **InstBenFG** holds (cf. p. 352 above). But, as we said, that principle is *not* at issue in the Function Argument.

And Aristotle finds the same teleological schema in parts of the body—and in organisms as wholes.

(ii) *Final Cause and Formal Cause*

The final cause is, as Aristotle remarks, in a sense the same, for these things, as their formal cause—the account of what they are (e.g., *Physics* II.7 198a24–6). The difference, as I understand it, is that the formal cause of something natural (or artificial) says what it is: so a knife is “cutting embodied in steel and wood,” the general formula of such essences being “such and such a Doing, in such and such

matter.” But with the formal cause the “such and such a Doing” is a first actuality—“cutting” in the sense of *the ability to cut*. For a knife doesn't stop being a knife when it is no longer actually cutting, and a human doesn't stop being a human when not actually exercising human life-activities, as when asleep. By contrast, the final cause is second actuality—the exercise, or fullest actualization, of essence.

It is obviously very important that the end of a knife is cutting as a *second actuality*. For if its end—the principle of organization of all things knife-ly—had been cutting in the sense of a first actuality, an ability or disposition to cut, then things would be very different: for, given *that* end, it would, say, be good for a knife to have an unopenable sheath, so that its ability to cut couldn't be damaged.<sup>37</sup>

### (iii) *The Final Good and the FG Inference Principle*

Now if it is “the human good” in this *teleological* sense of “the good of man” that the Function Argument clarifies, then this affects how we need to understand the inference principle **FG**. **FG** gives the rationale of looking to human function to determine human good. And so we need now to understand it not as [**BenFG**], but as [**TelFG**]:

[**TelFG**] Where the X is something with a function, the X-an good, i.e., *the good of an X*, consists in its doing its function successfully or well.

In types of functional thing, their good—their end and their organizing principle—is constituted by their function: actually cutting well is the good of—or success (*to eu*) in—the knife. So if the human has a function we can get a bead on its good from a consideration of its function.

On the teleological understanding of the X-an good, the **FG** principle itself seems, *pace* Glassen (p. 322), unpuzzling. (Admittedly there are still questions about the exact conditions for something's having an *ergon*: e.g., does a thief as such have an *ergon*?). Indeed **TelFG** has two advantages over **BenFG**. First it offers an intuitively self-evident principle about functional items—that with them their good or success lies in their function, in their actually functioning successfully. Such obviousness is needed given **FG** is supposed to provide the argument's rationale. Secondly, **TelFG** holds unrestrictedly of functional things—of knives, and eyes etc. And Aristotle, read at face value, is not at all concerned to restrict the **FG** principle—indeed the flow of his presentation is towards complete generality: “flute-player. . . every artisan. . . and generally everything of which there is some function and action.”<sup>38</sup>

In short, if we take **P2** as expressing **TelFG**, and understand “the human good” in the argument's conclusion teleologically, as “the good of man,” then Aristotle's argument seems valid, and sound—at least to the extent that it is plausible to view humans as having a function, and the one he claims they have.

## 20.6 Conclusion

As I understand his approach, Aristotle would, where *X* is a functional item, deny that any notion of *X*'s interests or welfare could intelligibly come apart from a notion of *X*'s function or end—its functioning successfully. In the peculiar case of humans it is part of our end, of our proper functioning, to deliberate and work out what constitutes our proper functioning, our acting correctly and successfully, in general and in particular. But there is no room at this level for an opposition between, say, prudence and morality, where the one (whichever) is viewed as our proper functioning and the other, by contrast, as our *real* interest or greatest good. This doesn't mean that tensions between considerations of self-interest and justice can't, or don't, arise *in* our working out what counts as acting well, and that humans can't, and don't, misconceive how best to act, and do so in selfish ways, etc. Aristotle's approach here doesn't prevent such disputes or tensions, but is rather one about *their conceptual, or categorical, location* (cf. Lawrence 2006).

In short, contrary to Wilkes' remark that

it is far from clear. . . how. . . the superb functioning of any *ergon*-bearing creature is relevant to what that creature's greatest good is

I take it that it is precisely Aristotle's position that with any *ergon*-bearing thing their greatest good couldn't be anything but their excellent functioning (and if unqualified, then such excellent functioning in circumstances and situations that are optimal in the case of the thing at issue: cf. Lawrence 1993 *passim*; see note 8 below).

More defense is required. But I hope some of the initial ground has been cleared.

**Acknowledgments** I first read Gerasimos Santas' work as a student in the early 1970s. Then as now I find his work on Plato and Aristotle admirably provoking and liberating. I am delighted and honored to dedicate this paper to Gerasimos. Its gist was presented at the APA in 2000. I thank my commentator Gabriel Richardson Lear and the Chair, John Cooper. I should also like to thank the editor Georgios Anagnostopoulos, and the colloquium at UC Davis, for their help. Special thanks are due to Julie Tannenbaum.

## Notes

1. The long, somewhat anacoluthic, sentence (1098a7–18) has already reprised the point that the *energeia psuchês* specific to humans is broadly rational—*kata logon ê mê aneu logou* (1098a7–8)—and is a matter of rational actions (*praxeis meta logou*: 1098a13–14). And so, I take it, Aristotle feels he can leave to the reader to supply this in the conclusion at a16.
2. Plato is an obvious influence on Aristotle here, and attracts comparable charges of fallacious inference. For example, Burnet, in his note on Aristotle's claim in I.4 1095a19 that people judge *to eu prattein* (and *to eu zên*) to be the same thing as *eudaimonia*, refers to Plato *Alcibiades* I 116b:

Whoever acts finely [*kalôs*] does he not act well?—Yes.—And those who act well are they not *eudaimones*?—Yes, how could they not be?

Burnet comments on the “ambiguity of this phrase [*sc. to eu prattein*]” that:

We must remember that Aristotle is here giving the views of others and *is not answerable for the fallacy*. He himself gives a very complete proof below 1098a7ff. (p. 15)

However it is not clear to me why Burnet thinks Aristotle's "complete proof" avoids the alleged fallacy.

The passage from the *Alcibiades* is not alone. In other passages too Plato accepts the equivalence, or identity, of:

- (1) living rightly/justly (*dikaiôs*) and living finely (*kalôs*) or living well (*eu*);
- (2) living well (*eu*) and living happily (*makariôs/eudaimônôs*).

See e.g., *Crito* 48b8–9. Of special importance is the precursor of Aristotle's own function argument at *Republic* I. 353e–54a. On its influence see further Lawrence (2001, n. 11). For charges of fallacy there, see Ackrill (1973, p. 20) and Irwin (1995, p. 179)—with neither of whom I agree.

3. Though I use "=" for convenience, it is more illuminating to think in terms of an "is" of constitution than one of identity. With the latter, we have an equation, and there is then a question of which way to read it (cf. McDowell 1980, sections 12–13). With the "is" of constitution that question doesn't arise.
4. This is charitable. It is not clear how far Wilkes or Whiting really appreciate that it is B, not A, that is at issue. (Thus Whiting accepts Ackrill's description of the inference to be defended as being "the move from . . . what it is to be a *good* man to . . . what is *good* for a man" (1988, pp. 34, 36): this sounds like  $2 = A$ , and so, in my view, involves a double misrepresentation.)
5. I take the noun understood with "*hekaston*" as "function" (cf. Reeve 1992, p. 127). If so, Aristotle is *here* saying that an excellence is *oikeia* to an activity/*ergon*, *x-ing*, rather than to its possessor X, as such. But Aristotle clearly holds the latter as well: cf. *NE* II.6 1106a15–18.
6. Aristotle assumes the equivalence of "*agathos* X" (good X) with "excellent X," i.e., an X with the excellence(s), *aretê*, of an X, and of doing something "successfully or well" (*eu*), with doing it "excellently" or "in accord with the relevant or proper excellence" (*kata tèn oikeian aretên*).
7. For discussion of these points, and again of how substantial even the specification of the excellences is, see further Lawrence 2001, *passim*, and especially section 7.
8. Actually the point of I. 5 1095b31–1096a2 is that the *possession* of virtue is not sufficient for *eudaimonia*; the worry *here* is whether even the *exercise* of virtue is always sufficient. This I call "the merely good life" problem: does a life of the exercise of human excellence always count as a *eudaimôn* one, however desperate the circumstances are under which virtue is exercised? If we think of *eudaimonia* in terms of success, of a person making a success of their life in the circumstances they are given, then I think we will see that "yes, a life of exercised virtue precisely constitutes making a success of one's life *in the circumstances*"—but this is compatible with one's having to *qualify* the claim that this "was a successful human life" by adding "given the trying and defective circumstances in which it was lived" (cf. Lawrence 1993; 2005, pp. 128–9): it is only a successful life to a secondary, or even more remote, degree. The unqualified claim would be false. The greatest humanly achievable good is an unqualifiedly successful life (although circumstances do not always allow of its realization).
9. Literally:

For just as in the case of flautist and sculptor and every craftsman, and in general of those of which there is some definite function [*ergon/product*] and actions, it is in their function that the good and the well/success is held to be, so it would be held also for a human, if at least there is some definite function of it.

10. Does the sculptor's good lie in his sculpting well, or in his sculptures being good ones? One could see the phrase "*ergon ti kai praxis*" as pointed—whether as "product and action" or as "function, i.e., action." But I think Aristotle is not concerned here with the difference (cf. *NE* I.1 1094a3–6, a16–18). We are to think primarily, I think, of the action; of course with producings, whether they are successful depends at least in part on whether they result (or tend to result) in good products. (Aristotle later uses the example of the lyre-player perhaps because it is easier to side-step this issue of separate product.)



11. If he does so, will he be a good sculptor? Yes, if he has achieved this end “sculpturally” and not by chance or at the instruction of another: that is, if his work is not merely in accord with the relevant excellences, but done with them, out of his own possession of them (*NE* II.4 1105a21–26; there are further conditions on possessing excellences of character, 1105a26–33).
12. I believe we already go wrong in supposing that there could, when thought through, really be a gap either between the human good and living successfully as a human, or between living successfully and living excellently. (i) The human good consists in living as it is human to live—actually doing what it is human to do—and doing it successfully; (ii) doing it successfully consists in doing it excellently—where the relevant excellences are those pertinent to the activity at issue, and so which constitute its (material) criteria of success, as also the traits of a success as a member of the kind. Now (i), if not exactly conceptual, seems a clarificatory thesis about the correct logical category of answer: thinking that the human good is, for example, a matter of being a good human is an error of logical category (cf. Lawrence 2001). The other, (ii), seems virtually a conceptual connection, given no substantial view of the excellences is being presupposed. Yet of course not just anything could be said to count as living successfully or excellently and be understood (cf. *Pol* VII.1 1323a27–29): these are principled concepts. How for instance (*pace* Erasmus’ joke) could folly, or practical stupidity, be viewed as a human excellence? This doesn’t mean that there mightn’t be great difficulty in working out what was wise or courageous or just to do. These are topics for discussion elsewhere.
13. McLaughlin (2002) remarks:

It would be easy, Aristotle tells us in the *Nicomachean Ethics*, to determine what a good man is if we only knew what the function or characteristic activity (*ergon*) of a man is. A good pruning knife is one that performs its function well and a good man would be one who performs his function well. (p. 140)

Herein he appears mistakenly to suppose that the object of Aristotle’s Function Argument is to define *what a good man is*, or *goodness* in a man. Glassen at least doesn’t suppose that that is Aristotle’s purpose. Again it is unclear whether McLaughlin is alive to the difference between first and second actuality ambiguities in claims of the form:

*The good X is one that performs its function well.*

These are between

- (i) the good X is one *that is so disposed as to x well*, i.e., to perform its function well, and
- (ii) the good X is one *that is actually x-ing well*, i.e., is actually performing its function well.

If the claim is supposed to be a *constitutive* one about what being a good X amounts to, (ii) would be false, in that a knife removed from actual chopping to the drawer would thereby cease to be a good knife. It is true that good knives do actually cut well—at least if certain conditions, e.g., about proper use, are observed: that is, a claim about goodness in an X has conceptual implications about what to expect in actual performance, under normal conditions, or absent interference, etc. So (iii) the good X is *one that actually x’s well*, i.e., it actually performs its function well, if and when put to use.

But this is rather a criterion of being a good knife, not a constitutive claim. More needs to be said about the relation of explicitly dispositional statements and such conditionals as (iii) (e.g., is (iii) an analysis of a disposition? etc.).

14. We could perhaps modify Glassen’s position and suggest that Aristotle in effect equivocates on **P2**. We could concede that in **P2** Aristotle is asserting **FG**, but then claim that in **FG** itself Aristotle is equivocating on the ambiguity of “the good”—asserting it in the “final end” sense of “the good” needed to give the argument’s rationale and to license the inference to the argument’s actual conclusion, **H3**, while confusedly relying on the substantial, “goodness of,” sense to secure the principle’s obviousness. If so, this is not far from the original accusation of fallacious inference.

15. And in Latin (e.g., Aquinas, *Summa Theologiae*, Ia2ae, Q 55, third reply; Q 56, first reply).
16. The difference between capacity (*dunamis*) and state (*hexis*) doesn't matter in this connection (cf. *EE* II.1 1218b38).
17. So too when she says "'good knives cut well' must be held to be some kind of analytic statement" (p. 135), she is, I take it, using "cut well" in the first actuality sense of "able to cut well."
18. Aristotle takes second actuality to be primary. (It is certainly so grammatically: for the first actuality is a disposition or ability with respect to second actuality.)
19. Glassen is not alone. Ackrill makes a similar mistake when, in accusing Aristotle of a slide, he says: "it is not self-evident that the best possible thing for a man is to *be the best possible man*" (p. 20). As Aristotle would quickly point out, it *is* self-evident that this *cannot* be the case: for "being the best possible man" is a dispositional notion, compatible with being asleep, or undergoing great torture. Ackrill's later formulation avoids this mistake:

It is not self-evident that the most desirable and enviable life for a man to lead is the life a good man leads. (p. 244)

Nonetheless this later formulation is preceded by the remark

However, even if such a "function argument" can establish something (very general) about the criteria for being *a good man*, it is not clear that it thereby establishes anything about *the good for man*.

He supposes, confusedly, that the argument is in part out to establish something about the criteria for being a good X (a good man). (Neither Wilkes nor Whiting are clear on these issues, as is evident from the variation in their formulations and from the fact that both take Wilkes and Glassen to be making the same objection.)

20. In effect Whiting distinguishes instrumental and beneficial senses of "*good for y*"—where in the first case *y* is an end or activity that is instrumentally promoted, and in the second *y* is someone, or some organism, object that is benefited (cf. McLaughlin 2002, p. 131).
21. In Whiting's description the moves are:

...commentators have often viewed his argument as consisting of two moves—first, the move from (a) what it is to be a *man* (or the function of a man) to (b) what it is to be a *good man*; and second, the move from (b) what it is to be a *good man* to (c) what is *good for a man*. (p. 34)

She *apparently* supposes, like Glassen, that the point of **P2** concerns a move from an X to a good X: being a good X, or *goodness* in an X, lies in X's function—for it is a matter of being *able* to do that well. In this vein she remarks:

[Aristotle] argues that the good and the (doing) well (*tagathon kai to eu*) of a flute-player or a sculptor or of anything with which has a function is determined by that thing's function. ... *A good flute-player* has the *virtue or ability* which enables him to perform well; a good knife is sharp and *able* to cut well. (p. 33, my emphases.)

22. Although Whiting confusedly switches from one formulation to another, this formulation of the objection accords with that on p. 35.
23. Cf. Ackrill (1973, p. 20):

The danger of the slide is not of course apparent in the case of a knife or an eye, since we do not raise questions about the welfare of a knife or an eye, or regard them as deriving benefit from their performances.

24. The point at issue in *NE* VIII.2 is that there is no friendship with inanimate things (a) because there is no reciprocity of friendship and (b) because there is no wishing for good things *for/in* the case of an inanimate object (e.g., wine)—unless merely as an intermediate beneficiary, as

one might wish for a wine's survival or its developing a good bouquet, only ultimately in order that one should enjoy drinking it: by contrast, it is a criterion of a friend that one "should wish for good things for his sake" (1155b31). (There are questions also in this passage over whether the datives (*ekeinô(i) agathou*) are of one's advantage ("good for") or of respect ("good in the case of").)

25. "The potential benefits of playing the flute depend on what desires and interests I happen to have"—and so are "relative and conditional," as against things that depend on my essential nature (p. 36). In fact Whiting seems to be to be operating with two distinctions here. One is in effect between the subjective and the objective—between benefits that are conditional on desires and interests I happen to have, and benefits that are "unconditional" or conditional on the essential nature of the natural kind at issue. The second is between instrumental and non-instrumental benefits (cf. things that are "intrinsically and non-instrumentally beneficial" for it, p. 37).
26. The *NE* appeals to "the flute-player and sculptor and every craftsman, and generally of those things of which there is some function and action" (1097b25–6); to the carpenter, cobbler (1097b28–9), and lyre-player (1098a9ff); and to organs of the body (1097b30ff). It does not explicitly mention artifacts, but these may be included in "those things" at b26 (see next note). By contrast, the *EE* II.1 appeals explicitly (a) to the function of artifacts: the functions of a cloak, a boat, a house and other things (1219a2–5); (b) to the function of skills—house-building, medicine (a13–15); cobbling (a20f); (c) of seeing, and of mathematical science (a16–17). The *Republic* instances (a) horses; (b) organs of the body—eyes and ears (352e5ff); (c) artifacts, pruning hook (353a4).
27. Aristotle gives examples of craftsmen, and then generalizes, first to every craftsman, and then further (*holôs*) to any case where X is "one of those of which/whom there is some function and action" (1097b26). The extent of this last extension depends on whether "*hôn*" is masculine or neuter. In favor of the masculine is perhaps the fact that Aristotle's interest in "*praxis*" so far in *NE* I has been with human rational action (covering technical as well as chosen action). In favor of the neuter are the following. (i) Gauthier/Jolif (Vol. II.1, p. 55) take it as neuter, comparing Plato *Rep.* 353a and *EE* II.1 1219a19–20. (ii) If it were masculine, the generalization cannot be that over artisans—for that generalization has already been accomplished by the preceding clause (*kai panti technitê(i)*). So one would have to assume that Aristotle was interested here in some even more general notion, say, of a role, or perhaps of knacks as well as skills strictly speaking (cf. Plato's *Gorgias*). But I haven't yet found a parallel in Aristotle for such an interest. (iii) On the neuter reading, the extension would naturally include not only artifacts like the infamous knife, but also body parts—the latter being examples he himself goes on to adduce a few lines later.

Whiting (her n. 4) toys with the idea that the absence of artifactual examples from the *NE* version is deliberate, thus negating the parallels that Gauthier/Jolif draw. But, first, I think that, *given* Plato's use of the pruning hook, and Aristotle's own use of artifactual examples in the *EE*, were Aristotle to have changed his mind, and considered that **FG** did *not* apply to artifacts, he would have needed to say this explicitly—something easy enough to say. And, second, there is the case of body parts.

But in any case the explicit parallel with the craftsmen suffices for the point against Whiting (although for an attempted emendation, see n. 29).

28. Whiting doesn't seem fully aware of it; or rather to deploy it in a different way (influenced by Cooper): see her n. 13. Wilkes' position is more complex. She thinks the inference from what the dog *qua* dog does to what is good for it does succeed (because of some feedback); but that the inference from what the good sheepdog does to what is good for the sheepdog does not. But, if I follow her argument, this relies on our understanding "good for the sheepdog" as "good for the sheepdog *qua* dog."
29. Kenny's clarification suggests the possibility of emending Whiting's position, as follows.
  - (i) Concede that **P2** expresses **BenFG**, but adopt the narrow view of "*hôn*," thus limiting the scope of **BenFG** to natural kinds and artisans and quasi-artisans.

- (ii) Still accept Whiting's way of side-lining artifacts.
- (iii) Given Kenny's *qua* point, deal with the artisanal counterexamples by extending Whiting's notion of categorical benefits to artisans and quasi-artisans. So playing the flute well will be good for the flautist *as such*, quite independently of what interests or desires he has as a human being.

But (a) this still asks too much of the reader, and (b) the case of body-parts remains unaccounted for. (See also n. 27 on "*hôn*".) Moreover, as I now go on to argue, (c) the claim that flute-playing is good for the flautist as such is itself not unproblematic.

- 30. Of course it may be the source of much satisfaction, pride, and income to the doctor as a human being. And practicing medicine may benefit the doctor in the sense of honing and improving his medical knowledge and expertise, his skill: but that is to consider imperfect doctors.
- 31. Earlier I put the ambiguity as one over *genitives* (see p. 336 above). If we take "the human good" as equivalent to "man's good," or "the good of man" then this might either be "what is good for, or of benefit to, a human" or "the good that belongs to a human as such"—a sense that I intend by "the good *of* man."
- 32. Cf. Aristotle is making the claim

...that the ways in which a thing can be benefited are at least partly determined by the kind of thing it is and what its essential properties are. (p. 38)

Second he must defend the alleged connection between the essence of a kind and what benefits members of that kind. (p. 40)

- 33. Whiting conceives of Aristotelian essentialism as normative all the way down (ibid. pp. 38–9).

Aristotle does not think we can give an account of the essence or the function of a kind without introducing some notion of what is beneficial for members of that kind. This is what I had in mind when I said that Aristotle's account was normative 'all the way down.'

In a sense—though one that needs considerable explication—I agree that nature is normative all the way down, but I do not understand why Whiting supposes the normative should be explicated in terms of the beneficial. Certainly there is a good in question here, but I believe it is the good of man, not what is good for man.

- 34. The question "what's the good of it?" is itself ambiguous. It can be understood in the sense I had in mind, as asking after the use, or point, of something (cf. "what does it Do?"). But in context it can be used to ask "what's the good of it *to us*, in our present need?"—i.e., "what use is it to us?" or "what help, or advantage, does it offer us?" But even this is to ask what benefit it confers, and not what benefits it.
- 35. Aristotle adverts to this double sense of *hou heneka* several times: *An* II.4 415b2, b20; *Met*  $\Lambda$ .7 1072b1–3; *EE* I.8 1218b10 (where "*hos telos*" is added to "*to hou heneka*" to disambiguate it—a point Woods fails to register in his translation (1982); cf. *Met*  $\alpha$ .2 994b9), 8.3 1249b12–16; *Phys* II.2 194a35–6, with its reference to the *De Philosophia*. (The *Phys* II.2 passage may suggest that the beneficiary too can in a certain sense (*pôs*) also be called an end (*telos*)). Hicks commenting on the *An ad loc* (p. 340) refers also to *GA* II.6 742a22. Christ too quotes it. But Ross correctly says it is not a parallel (ad *Met*  $\Lambda$ .7 1072b1–3, Vol. II, pp. 376–7).
- 36. That is, the X-an good could be either "that which is for the sake of Xs"—i.e., what is good for, or benefits Xs, or else 'that for the sake of which Xs are,' in the sense of the point or end of Xs.
- 37. In fact it is not clear whether, when pushed, this thought holds up. The question is whether the final cause of some thing *could* intelligibly be a first actuality; or whether a first actuality is in a way conceptually dependent on a second actuality. If so, then in the above case with the

knife, it would have to have a kind of actual use—"hanging around being able to cut"—which was its second actuality, and its principle of organization. I suspect that this is so, but will not pursue it.

38. (1) As should be clear by now, I see no reason to suppose that Aristotle is excluding either body parts or artifacts from the class of "what generally has some function and action."  
 (2) Any hesitancy he has over his strategy is directed rather to whether humans have a function (*eiper esti ti ergon autou; tacha dē genoit' an. . .*), and so whether the human good that is the end of our rational actions will be illuminated this way. It is the application of the principle to the humans that might be doubted, not the principle itself.

## Bibliography

- Ackrill, J., ed. 1973. *Aristotle's Ethics: Selections from Philosophers*. London: Faber.
- Austin J.L. (written pre-1939/1967). "Agathon and Eudaimonia in the Ethics of Aristotle." In *Aristotle: A Collection of Critical Essays*, edited by J.M.E. Moravcsik. London: MacMillan.
- Barnes, J., M. Schofield, and R. Sorabji, eds. 1977. *Articles on Aristotle: Vol. 2*. London: Duckworth.
- Burnet, J. 1900. *The Ethics of Aristotl*. (Edited with introduction and notes). London: Methuen.
- Crisp, R., ed. 1996. *How Should One Live?* Oxford: Oxford University Press.
- Foot, P., ed. 1967. *Theories of Ethics*. Oxford: Oxford University Press.
- Foot, P. 1978. "Goodness and Choice." In *Virtues and Vices*. Oxford: Blackwell Publishers. (Reprinted from *Aristotelian Society Supplementary Volume 35*, 1961, 45–60)
- Geach, P. 1956. "Good and Evil." *Analysis* 17:33–42. (Reprinted with alterations in P. Foot 1967.)
- Glassen, P. 1957. "A Fallacy in Aristotle's Argument About the Good." *Philosophical Quarterly* 7:319–22.
- Hooker, B. 1996. "Does Moral Virtue Constitute a Benefit to the Agent?" In edited by R. Crisp.
- Hurka, T. 1987. "'Good' and 'Good For'." *Mind* XCVI:71–3.
- Irwin, T. 1995. *Plato's Ethics*. Oxford: Oxford University Press.
- Joachim, H.H. 1951. *Aristotle, The Nicomachean Ethics*. Oxford: Clarendon Press.
- Kenny, A. 1965–1966. "Aristotle on Happiness." *Proceedings of the Aristotelian Society* 66:93–102. (Revised and reprinted in J. Barnes et al. 1977).
- Lawrence, G. 1993. "Aristotle and the Ideal Life." *Philosophical Review* 102(1):1–34.
- Lawrence, G. 2001. "The Function of the Function Argument." *Ancient Philosophy* 21:445–75.
- Lawrence, G. 2005. "Snakes in Paradise: Problems in the Ideal Life." *Southern Journal of Philosophy* 43:126–65.
- Lawrence, G. 2006. "Human Good and Human Function." In *The Blackwell Guide to Aristotle's Nicomachean Ethics*, edited by R. Kraut, 37–75. Oxford: Blackwell.
- McDowell, J. 1980. "The Role of *eudaimonia* in Aristotle's Ethics." *The Proceedings of the African Classical Associations XV*. (Reprinted in A. Rorty 1980).
- McLaughlin, P. 2002. "On Having a Function and Having a Good." *Analyse & Kritik* 24:130–43. Stuttgart: Lucius & Lucius.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Prichard, H.A. 1935/1949. "The Meaning of *agathon* in the Ethics of Aristotle." Reprinted in Prichard, H.A., *Moral Obligation*. Oxford: Clarendon Press.
- Reeve, C.D.C. 1992. *Practices of Reason*. Oxford: Clarendon Press.
- Rorty, A., ed. 1980. *Essays on Aristotle's Ethics*. Berkeley/Los Angeles: University of California Press.
- Thompson, J.J. 1994. "Goodness and Utilitarianism." *Proceedings of the American Philosophical Association*, Presidential Address, 67(4):7–21.
- Urmson, J.O. 1988. *Aristotle's Ethics*. Oxford: Blackwell Publishers.
- Vendler, Z. 1963. "The Grammar of Goodness." *Philosophical Review* LXXII:446–65.
- Von Wright, G.H. 1963. *The Varieties of Goodness*. London: Routledge & Kegan Paul.

- Whiting, J. 1988. "Aristotle's Function Argument: A Defense." *Ancient Philosophy* 8:33–48.
- Wilkes, K. 1978. "The Good Man and the Good for Man in Aristotle's Ethics." *Mind* 87:553–71.  
(Reprinted in A. Rorty 1980).
- Woods, M. 1982. *Aristotle's Eudemian Ethics*. Oxford: Clarendon Press.
- Ziff, P. 1960. *Semantic Analysis*. Ithaca, NY: Cornell University Press.

# Chapter 21

## Aristotle on Discovering and Desiring the Real Good

Mariana Anagnostopoulos

### 21.1 Introduction

Aristotle's treatment of the type of desire he terms "wish" appears to be a rejection of the Socratic idea that the desire for the real good is universal, as it is the ultimate desire of virtuous and vicious alike. Indeed, in Plato's Socratic arguments, the ultimate good and good things are the only objects of desire. At *Nicomachean Ethics* III.4 Aristotle considers the important question whether we desire what appears good, what is really good, or both, and concludes that an accurate account must include both objects. In denying that wish is exclusively for what is *truly* good, Aristotle would seem to commit himself in some way to the alternative, that wish is for what *appears* good. His conclusion, however, is that the object of wish is the apparent good only "to each," while "in truth" it is the *real* good. I will attempt in what follows to render meaningful the contrast Aristotle takes there to be between the two views he rejects, in a way that leaves room for the third option he endorses.

Penner and Rowe (1994, pp. 4–5) make a strong distinction between Aristotle and early Plato when they say that,

...we get, on the Aristotelian view... that everyone desires as their end their *apparent* happiness, not their *real* happiness. By contrast to this view, we claim that what the *Gorgias* is saying is that what one desires as one's end is one's *real* happiness, even if that differs from what one thinks it is.

Socrates, according to Penner and Rowe, holds that there are no conceptions mediating our desires for objects; despite our beliefs, we desire real objects—not the objects specified in our conceptions—and have these in mind when we act. It is only in mistaken cases, when an agent's beliefs are in large part false, that we must appeal to the agent's mistaken beliefs to explain actions toward objects that are bad. In these cases, they explain, the agent has an incoherent desire which does involve the agent's conception of the object, but is not clearly *for* the object that appears in that conception.

---

M. Anagnostopoulos (✉)  
California State University, Fresno, CA, USA  
e-mail: marianaa@csufresno.edu



Aristotle, on the other hand, is taken by Penner and Rowe to appeal to conceptions in a way that Socrates does not, and that Socrates would find objectionable. Because conceptions, or appearances, are involved in all cases of wishing in Aristotle's account, they argue, wish is for the apparent good on that account, even when the apparent good coincides with the real good. In his *Topics*, Aristotle cautions those who attempt to define wish without reference to conceptions: "You must see whether there is a failure to add the qualification 'apparent,' for example in the definition. . . 'wish is a desire for the good' . . . instead of 'the apparently good'" (*Topics* 146b35–147a2). The apparent good cannot be *the* object of wish, however, as long as there is such a thing as the good for humans, which is found in human well-functioning and is the end of all action *and desire*, as Aristotle contends. If his account of desire is to be compatible with other elements of his perfectionist conception of virtue, Aristotle must maintain that the real good is the ideal and ultimate object of wish. One contrast to Socrates is clear: for Socrates, desire is not for the apparent good at all. However, it is less obvious how Aristotle would treat the desire for the real good.

In considering which version of the idea that wish is for the apparent good we are justified in attributing to Aristotle, I will emphasize Aristotle's contentions that the real good is the only thing one should wish for, and that it is when one *is mistaken* that one wishes for the merely apparently good. The basis for these ideas is Aristotle's belief that the good person apprehends the world as it is; she succeeds, in desire, perception, and judgment, in identifying and apprehending what is true. So, like Socrates, Aristotle conceives of wish as fundamentally related to the *true* good. However, he departs from Socrates in attributing corresponding desires for things that are not truly good to those whose beliefs are mistaken. Socrates doubts that we desire whatever appears good, seemingly because he doubts that desire necessarily operates by way of conceptions of what is sought.<sup>1</sup> Without this reason to reject desires for merely apparently good things, Aristotle is free to do away with Socrates' distinction between what is desired and what is "thought best," and to posit a position distinct from that of the *Gorgias*. Although it is true that in Aristotle's account wishing is mediated by appearances of things, it is also clear that Aristotle takes there to be a proper and correct object of wish, the real good; in this he does not differ from Socrates and Plato.<sup>2</sup>

## 21.2 Aristotle's First Argument at *NE* III.4: Wish is Not Simply for the Real Good

Socrates explains in Plato's *Gorgias* that it is "in pursuit of the good" that we all act, and that we desire only those acts that turn out to be beneficial, or real means to the good, thus refuting the view—insisted upon by Polus—that desire is for whatever seems best to us (*Gorgias* 468b). Aristotle does not take this dispute to have been resolved when he turns his attention to *boulêsis* in the *Nicomachean Ethics* and notes, "that wish is for the end, we have already said; but to some it seems to be for

the good, whereas to others it seems to be for the *apparent* good” (*NE* 1113a15–16; translations as in Broadie and Rowe 2002, except where noted). As Aristotle considers the question he is exploring to be complex, it is useful for us to be able to identify relevant sub-questions. In his first argument, Aristotle finds that particular instances of desiring are left unaccounted for by the view that the *only* object of wish is the good. But in his second argument he identifies a proper, ideal, and best object of wish, and concludes that this is not the apparent good. We are left without a clear sense of whether he is seeking the proper or actual objects of this faculty.

When he first considers what I will call view “(S),” the Socratic view that “the object of wish is the good,” Aristotle finds that it is excessively restrictive and unable to account for all exercises of the faculty:<sup>3</sup>

**Argument (S)** The consequence, for those who say that the object of wish is the good, is that what the person making an incorrect choice wishes for is not wished for (for if it is wished for, it will also be good; but in fact it may have been bad. . . (*NE* 1113a16–19))

Aristotle’s argument seems to run as follows:

- (1) Wish is exclusively for the really good. (Assumption: View (S))
- (2) If  $x$  is an object of wish,  $x$  must be good. (From (1))
- (3) Agent  $A$  wishes for  $x$ .
- (4) Suppose  $x$  is bad, i.e. is a wrongly chosen end. (Assumption)
- (5)  $A$  doesn’t wish for  $x$ . (From (2), (4))

On the assumption, noted in premise (4), that one can be mistaken about which ends are good, (5) contradicts (3). The Socratic view has the implication—absurd, according to Aristotle—that one who makes a mistake in identifying her best end really doesn’t wish for that end. But it seemed as familiar a phenomenon to Aristotle as it does to us that desire and action often follow upon false beliefs about what is good to do.<sup>4</sup> If the assumption in (4) is incontestable, it must be said about the misguided person that, “what he wishes is not wished, which is self-contradictory,” as Irwin notes (Irwin 1999, p. 36).

A second implication of view (S) is not explicitly addressed in this passage: we cannot explain  $A$ ’s pursuit of  $x$  by appeal to  $A$ ’s desire for  $x$ . If we assume that  $A$  chooses  $x$  and acts on that choice, it follows from (5) that  $A$  acts toward  $x$  without any motivating wish for  $x$ , a wish that would otherwise provide the explanation for  $A$ ’s action. The lack of either a motivation on the agent’s part or an explanation of the agent’s action is unsatisfactory given Aristotle’s theory of desire and voluntary action, many components of which function as assumptions in argument (S). An examination of these background assumptions is necessary in order to expand this very terse argument, locate it within the larger framework of Aristotle’s theory of action, and show that if the object of wish is the real good alone, one who seems to wish for something bad doesn’t really wish for that thing and, for lack of a motivating desire, one’s pursuit of a bad end cannot be explained.

The mistaken ends possibility is instrumental in Aristotle's argument that wish is not simply for the real good *and* in his argument that wish is not simply for the apparent good. Aristotle insists that "the person making an incorrect choice" (*NE* 1113a17) is not a rare exception, but that those who correctly identify good things are few: "most people are deceived, and the deception seems to come about because of pleasure; for it appears a good thing when it is not" (*NE* 1113b1). Though Socrates would not deny that pleasure is attractive, he would characterize our misguided agent as erring with respect to belief. The mistaken apprehension of a harmful pleasure as good does not imply a corresponding motivating desire, according to Socrates. Though it is tempting to think Socrates must have held otherwise, I follow Penner and Rowe (1994) in maintaining that the two thinkers diverge on this point. Aristotle parts with Socrates in rejecting what Penner and Rowe (1994, p. 18) identify as the main tenet of Socratic doctrine: "people do not differ ethically in their desires, wills, or characters, but only in their knowledge."

In order to understand the dispute about the mistaken cases, let us consider the specific ways in which error can occur in an agent's deliberation about how to act. In the *Eudemian Ethics*, Aristotle observes that an agent may go wrong in relation to means and ends in action in several ways: "It is possible to have one's aim right but to be entirely wrong in one's means to the end aimed at; and it is possible for the aim to have been wrongly chosen but the means conducing it to be right; and for neither to be right" (*EE* 1227b20). Since it has been established that "wishes... are for ends," if it is possible that one's end be wrongly chosen, wish in such a case will be for something other than the real good. Socrates would agree only that *beliefs* might be mistaken in these ways, but not desires. It must be that Aristotle relies on an assumption Socrates does not share: that choice leads to desire for the object of choice. With an ultimate goal (*eudaimonia*) in place, other ends are chosen as intermediate to this. The contention is that these can surely be chosen poorly, as can the nature of happiness be misunderstood.

To summarize the two problems with respect to belief, desire, and action: As Joachim (1962, p. 104) remarks, on the view that wish is for the real good, we are "forced to deny that people with mistaken ideals desire at all." Secondly, and contrary to what is generally believed, we must say that bad or misguided actions are not guided by a motivating desire for the very action pursued. If we take into account Aristotle's view that desire, either for particular things or for ends, plays a primary causal role in determining action, it becomes clear why he cannot accept the view that wish is simply for the real good.<sup>5</sup> In the *De Anima*, for example, Aristotle considers whether desire and reason are distinct motivating forces in the soul, and concludes that there are not two things, but only "one thing which produces movement, the faculty of desire."<sup>6</sup>

Reason alone will not explain the pursuit of wrongly chosen (or any other) ends. If there can be an agent like the one in argument (S)—one who does not wish for the bad end he pursues—there must be some other desire motivating the agent's action. From Aristotle's assumption of the mistaken ends possibility, taken along with certain of his other convictions, we may formulate a second forceful argument against the view that wish is only for the real good. Aristotle states in the *Eudemian*

*Ethics* that “nobody wishes what he thinks to be bad” (*EE* 1223b8). These claims may be combined into an argument that Aristotle doesn’t formulate explicitly but that I think he would accept:

- (1) In acting toward end  $x$ , agent  $A$  must be motivated by some wish.
- (2) There can be actions toward wrongly chosen (bad) ends.
- (3) Agent  $A$  does not wish for any end  $x$  thinking  $x$  is bad. It then follows that:
- (4) For any end  $x$ ,  $A$  wishes for  $x$  thinking  $x$  is good.

In the case under consideration,  $A$ ’s wish must be for what is (merely) apparently good.

Thus, two things speak against view (S): (1) its contradictory implication that what is wished is not wished, and (2) the fact that desire for what appears good (but isn’t) is needed for explanations of some actions. It is evident, both from the objection Aristotle makes at *NE* 1113a22 and from the one that can be drawn from his comments in the *De Anima* concerning the role of desire as motivating action, that he rejects the Socratic postulation of a single motivating desire for the real good. When Aristotle insists, in the *Topics* passage noted above, that wish must be defined with reference to the apparent good, he seems to advance argument (S) by stating explicitly his alternative to Socrates’ mistaken view. We learn that an adequate definition of wish must cover all instances of the exercise of the faculty, which view (S) does not do.<sup>7</sup>

In his defense of the qualified definition, Aristotle’s reasoning is almost identical to that of argument (S): what is wrong with the definition “*wish is a desire for the good*” is that “often those who feel desire fail to perceive what is good or pleasant, so that the object of their desire is not necessarily good or pleasant, but only apparently so. One ought therefore, to have assigned the definition with this qualification” (*Topics* 147a2–5). We have found that the apparent good must appear within a correct definition of wish. It would be truly surprising, however, if Aristotle opted to define wish exclusively with reference to this object, for this seems to be exactly the position he rejects in the second part of his analysis at *NE* III.4.

Aristotle ultimately characterizes the virtuous person as one who participates, via his or her practical wisdom, in a unique relationship with the real good, and does not merely encounter it by accident. I take the mistaken ends possibility to reinforce just this part of the theory. To say that desire for the apparent good involves some kind of mistake—not merely in *determining* the good, but in *desiring it*—is to say that rational desire can and should be characterized with reference to something other than the apparent good.

We have a hint of this in another part of the *Topics* passage, one that is perplexing when considered alongside the segment just discussed. There, Aristotle notes the requirement that a definition of a relative term specify that to which the term is relative. Wish, therefore, must be described as “desire of the good” (*Topics* 146b5). We encounter a problem: argument (S), as we have seen, reduces this definition of wish to absurdity. It is also inconsistent with the second faulty definition, given just a few lines below it in *NE* III.4: that wish is for the apparent good. Turning, in the

next section, to Aristotle's argument (**P**), that wish is not *properly* for the apparent good, we will see why neither of the claims of the *Topics* passage constitutes a strict definition of wish but instead both pertain to components of a complete analysis of the faculty.

Thus, the *Topics* and *NE* passages are ultimately compatible once we consider the context of the claims therein. When Aristotle says that "wish is desire for the good," he means to illustrate his point about relative terms. *Boulêsis* is relative to good in the way that *epithumia* is relative to pleasure; in each case the faculty aims at some singular kind of thing. That wish is relative to the good is not to say that only the real good is desired, but that goodness is the feature targeted by desire. Wish is not for the bad, the pleasant, or the sweet but, very generally stated on Aristotle's part, for "the good." The view is given in such general terms elsewhere: at *Metaphysics* 1072a25 he argues that, "the primary objects of desire and thought are *the same*. For it is the apparent good that is the object of appetite, and the real good that is the primary object of wish [*boulêton de prôton to on kalon*]." Here is an example of Aristotle's use of "good" to capture the real good and the apparent good. Of course the good and the apparent good are *the same* only generally, in that they are or are conceived *under the aspect of good*. We should note two things from this passage: the *real* good is indicated as the *primary* object of wish and, unlike wish, appetite *does* take as its primary and proper object *the apparent good*. Wish, as the desire of the *reasoning* part of the soul, has its object determined by calculation and choice, unlike appetite. It is the job of reason to find out what is truly best for us to desire and do; as rational beings, we are suited to seeking, discovering, and living out a life in which we exercise our varied faculties well, in accordance with reason.

Similarly, when Aristotle says that wish must be defined with some reference to the apparent good, he aims to ensure that the resulting definition cover every case, including that of the good person for whom what seems good is actually good, and that of the rest, who wish for the (merely) apparently good due to flawed judgments about the real good. Though we must take seriously Aristotle's contention in argument (**S**) that people who judge poorly nevertheless desire what they take to be good, the ultimate aim of the passage is to identify the *proper* object of wish, that which reveals the function and character of the faculty in its most excellent condition.

### 21.3 Aristotle's Second Argument at *NE* III.4: Wish is Not Simply for the Apparent Good

The preceding discussion of the *Topics* provides a starting point for our examination of Aristotle's claim that wish is not properly for the apparent good. View (**P**), bearing similarities to certain ideas of Protagoras, provokes the following response from Aristotle:

**Argument (P)** And those on the other hand who say that what appears good is wished for, are forced to admit that there is no such thing as that which is by nature wished for, but that what each man

thinks to be good is wished for in his case; yet different, and it may be opposite, things appear good to different people (*NE* 1113a20–23, Rackham, transl.).

Both view (**P**) and Aristotle's objections to it, again very concisely stated, require clarification. Irwin notes an initial concern about the argument: “[its] exact point is obscured by the ambiguity of ‘wished’ between (a) what is wished and (b) what deserves to be wished.”<sup>8</sup> A point about (a) would make a descriptive claim such that people do in fact wish for the ends they perceive to be good, while a point about (b) would assert that the proper object of wish is the apparent good. In argument (**S**), Aristotle sought to clarify the *boulêton* by making a claim about (a), rejecting Socrates' view on the basis that it leaves out the apparent goods most people wish for *in fact*.

We now see that the aim in argument (**P**) is a different one—one that corresponds to other important remarks on correct definition in the *Topics*. Relative terms, of which “wish” and “knowledge” are given as examples, must be described in relation to something, namely an end that “is best. . . or ultimate,” an end being “in any particular case that which is best or that for the sake of which all else exists” (*Topics* 146b11–12). View (**P**) is inadequate because on that view, wish cannot be related to some best end; there is nothing among the apparent goods that is best or ultimate as an object of wish, if wish is merely for what appears good. “The paradox in this,” Evans explains, “consists in the fact that ‘the apparent good’ was given as an answer to a question about *the* object of wish; but as an answer it shows that the question was misconceived. For the question was about *the* object of wish, and the effect of the answer is to deny that there is such a thing” (Evans 1987, p. 56). This is why we must see Aristotle as concerned, in the two arguments, with two aspects of the initial question “what is the object of wish?”

Aristotle's inquiry begins with a single concern: to investigate the nature of wish, conceived abstractly. This problem is then sub-divided into two questions, the answers to which are required for a full account of the sort Aristotle is seeking: (1) what is the ideal or proper object of the faculty, and (2) what are all the (kinds of) objects of the faculty? These are not isolated questions; a theory defining a faculty must be compatible with and take into account evidence as to how the faculty operates. It should tell us what is desired, and also what we need to know to determine when what is desired is not what should be desired.

Considering either of the unacceptable views in isolation, the distinction between “*successful or proper exercises of a faculty and the mere exercise of the faculty. . . is obliterated*,” Evans explains. Like Irwin, Evans attributes some of the difficulty with the passage to the Greek word *boulêton*. Evans, however, understands the term not as ambiguous between two senses, but as a single concept that contains the two elements. He correctly points out that Aristotle's own view would not fare well as a response to both (**S**) and (**P**) were the object of his inquiry truly ambiguous between the two senses of the concept addressed separately in the rejected views (Evans 1987, pp. 60–61). It would, in that case, be open to the objection that it specifies the object of wish in one sense while ignoring the other. Thus, Aristotle must consider both aspects of the concept, both as he works through the faulty views and when

he ultimately provides his own answer to the question “what is the object of wish?” For example, if Aristotle intended simply to identify the proper object of wish, the possibility given in premise (4) of argument (S)— $x$  is a wrongly chosen end—would not result in a contradiction that what is wished is not wished. This contradiction results only if view (S)—that wish is for the real good—is understood to provide an exhaustive list of objects of wish. Similarly, if Aristotle were interested only in describing wish in a way that covered every case, view (P)—that wish is for the apparent good—would not be so problematic; it is insufficient as a full account because it fails to isolate excellent instances of wishing.

To complete our analysis of Aristotle’s second argument, it is helpful to look ahead to his own view, given in the remainder of *NE* III.4, to see what it contains that view (P) doesn’t. Irwin understands Aristotle’s own view to be that though each person wishes for what he takes to be good, the proper object of wish is exactly what “the well-informed person” wishes for, the real good. This indicates that the version of (P) that is rejected would have to be that wish can be characterized *solely* in terms of the apparent good as object; only this view entails that the *proper* object of wish is the apparent good. No contradiction results from supposing that wish has as its proper object the apparent good, but the consequence that there is “no natural object of wish” is unacceptable in a different way. We can better understand the term “natural” (*phusei*) by noting what Aristotle considers to be the undesirable alternatives to there being something wished by nature, and his use of the term in explaining what is involved in there being something *phusei boulêton*.

First, in the absence of a natural object of wish, what is wished will be (1) for each, what seems good to him, (2) different things, and (3) even contrary things. We can assume it is true that different people construe different and even opposite things as good, but Aristotle’s point must be evaluative rather than descriptive. If different and opposite things are *proper* objects of desire, there would be nothing that is *not appropriate* to desire. That contrary things may be wished for results in a relativity of a certain kind, if these are not ordered or ranked. Not only are opposite things desired; they are equally desirable. This means that the agent’s recognizing or misidentifying the good has no bearing on whether that person desires successfully. Though we do not speak of desire as true or false, we do identify a different criterion of success for desire. As will be seen later in the *NE* III.4 passage, on Aristotle’s conception of moral excellence, the good person must wish for the right end, the real good, recognizing it as the real good by virtue of his own goodness. If there is nothing wished “by nature,” there is no distinction to be made between appropriate and inappropriate objects of wish, in terms of what is *good*. In this case, the distinctions between, for example, the good, bad, weak, and continent persons, as Aristotle sees them, cannot be sustained. These states of character are *ranked* by Aristotle, partly on the basis of the desires they involve: for example, the weak-willed person might desire the wrong things or desire things in an excessive way, while the practice of aiming to maximize bodily pleasure is considered by Aristotle to be even worse than weakness of will, because it prevents the human being from developing his varied capabilities. This is something we would assume the virtuous person would understand; her judgment will reasonably be expected to influence her choices, wishes,



and actions, and *each* of these will therefore reflect the excellent condition of her soul.

Thus, the relativity entailed by (3) is the most important consideration against the view that wish is for the apparent good. Virtue and practical wisdom are what ensure that the good person wishes for the real good non-accidentally. Without such a requirement, the term “virtuous” can signify only an accidental coincidence of desire with goodness, and not a state of character that depends on and is characterized by desire for the good. This is why Grant calls the view that wish is for the apparent good “a corollary to the doctrine of Protagoras,” which amounts to the claim that “if the individual could only know what “seemed” to him, he could only wish for what seemed good” with the result that “the objective distinction between good and evil is done away with” (Grant 1973, p. 24).

Grant’s formulation of Aristotle’s concern brings out one way of reaching the conclusion that wish is for the apparent good. If one can only know one’s conceptions—without the possibility of checking these against what is the case—one cannot but wish for what seems good. Yet another way to defend a relativistic view of desire is to drop entirely the assumption that desire follows upon apprehension of goodness. One would subsequently be forced to deny that wish has any special tie to the good, and that the informed judgment of the virtuous person leads to that person’s success in desiring. We could not conceive of the good person as standard of goodness, since what a person who happens to hit upon the good judges in each case is no more fitting as an object of desire than what others judge healthy, good, or sweet.

This option will be hard to understand if our concept of rational desire is bound up with that of goodness: we might think that what it is to desire, rationally, is to seek to attain something beneficial or better. Herein lies the standard of good and bad desires; for desires based on beliefs, the error in desire can be traced back to a false belief about what is good. Grant’s reasoning (when he takes the absence of a *phusei boulêton* to signify a like absence of an objective good) must be that if we work with the standard concept of desire, and at the same time maintain that wish is just for the apparent good, there can be no mistakes in desire. This in turn must be *because* there is nothing truly good to serve as that standard. Joachim, like Grant, assumes that desire aims at goodness; the association of goodness with desire requires this, for otherwise “good” would be no more desired than anything else.

If it is possible to realize one’s actions have not led to one’s good, one must do more in desiring than aim at the apparently best option. To desire, one must perceive some object as good, but the apprehension of goodness does not occur independently of the faculty of understanding. Success in understanding will connect us to the true good, which will be the proper object of desire. This would not make sense, if it were no better to *want* the good than anything else. If informed judgment gets one further in identifying the good, one is better able to fill in one’s general desire for happiness with the things that will promote it. Is it not also peculiar to the real good that it is by nature desirable? It would be odd to endorse objectivity about the good and to hold simultaneously that desire does not aim at the good.

An analogous discussion of truth and falsity occurs in Book  $\Gamma$ , ch. 4, of Aristotle's *Metaphysics*. There he argues that, contrary to what Protagoras' doctrine that man is the measure of all things implies, it is impossible for the same thing both to be and not to be. If whatever appears to man truly is, all appearances are true or equal. But people usually regard beliefs opposite to their own as false, so these beliefs will be thought to be both true and false and will therefore be both true and false. Aristotle states, examining Protagoras' view, that "it is not clear which of these appearances are true and which are false, for the ones are no more true than the others, but both alike are true." And again, "all opinions must be true; for those who are mistaken and those who think truly have opposite opinions, and if this is the way things are, then all men think truly" (*Metaphysics* 1009b9–11 and 1009a8–16). To suppose that the true is whatever is taken to be true, and not what is discovered to be true, is to dismiss the definitive role of the real good in desire—the idea that what is truly good is truly desirable, though not always desired. If, on the other hand, "we may define a good thing as that which ought to be chosen for its own sake," it must be that the opposite thing is bad and should not be chosen for its own sake (or at all) (*Rhetoric* 1362a21). But we see on view (**P**) that the only contrast is between the apparent good and the apparent bad; this does not allow us to distinguish virtue and vice based on whether an agent wishes for the truly good or not.

A key difference between Socrates and Aristotle is that for Aristotle, virtue depends on more than knowledge; feelings, desires, and actions must all fall within a mean that benefits the individual (*NE* 1113a25–26). Penner and Rowe point out that, unlike Plato, Aristotle sees the difference between the virtuous and the rest as determined at least in part by what is wished for, and not merely by what is known: "for our apparent goods *do* differ. . . and make us differ in the quality of our desires" (Penner and Rowe 1994, p. 18). Some of these apparent goods, of course, are apparent goods only in that virtuous persons construe them as good; they are, in themselves, truly—not only apparently—good, and therefore, in construing them as good, virtuous persons understand them exactly as they are. Aristotle's claim with respect to the existence and apprehension of the good is that, "if not all things are relative, but there are some things which exist in virtue of themselves, not every appearance would be true" and "clearly those on one side must be mistaken" (*Metaphysics* 1011a17–21 and *NE* 1062b35).

## 21.4 Aristotle's View: Wish is for the Good or the Apparent Good

It is with Protagoras' view in mind that we return to our main text, and find that Aristotle has more to say about appearances and truth. First notice the conclusion Aristotle draws with respect to wish,

But if, then, we are not content with these views, should we say that the good is without qualification and in truth the object of wish, whereas what appears good to a given person is the object of wish for that person?

### 21.4.1 *The Absolutely Good (haplôs agathon)*

The real good is the sole object that is naturally (*phusei*) and absolutely (*haplôs*) an object of wish. Aristotle comments in the *Politics* on his own use of the term in the *Ethics*, noting that he had “called absolutely good what is good in itself” (*Politics* 1332a10). It will turn out that the reason the real good is wished *haplôs*, according to Aristotle, is that it is good *haplôs*. Something is absolutely good, in Aristotle’s view, if it is good simply rather than good merely to a particular person: “some things are. . . good absolutely whereas others are so to a particular person but absolutely are not so, but on the contrary are bad,” for example “all the things which are beneficial to the base” (*NE* 1152b27, 1235b32, 1228b18–21). An ambiguity that occurs here and elsewhere in the *Nicomachean Ethics* is directly relevant to our inquiry; what is good relative to a particular person is either what is good *for* a person or what is good *according to* a person. The same problem can be formulated about the absolutely *desirable*.

In fact, a parallel can be drawn between the lovable and the wishable that illuminates Aristotle’s view of the *haplôs boulêton*. About the lovable, Aristotle says: “It is agreed that what is good and pleasant absolutely (*haplôs*) is lovable and desirable strictly, while what is good and pleasant for a particular person is lovable and desirable relatively to that person” (*NE* 1157b27). Here, we also need to know whether Aristotle has in mind what is lovable for a given person or according to that person. In addition, we see again the two questions about the objects of a faculty: in stating what is lovable or wishable absolutely, we answer the question, “What is properly or ideally lovable or wishable?” while the question, “What do people love or wish for?” may be answered with reference to how objects of love or wish appear to those who desire them.

Aristotle answers the first question and explains that the relationship between the good relative to a person and the good that is absolutely so is that, “what is sought for is that things absolutely good shall be good for oneself. For the absolutely good is absolutely desirable” (*EE* 1237a1). One aims in one’s life to desire and attain the things that are truly good; success in this endeavor will mean that the things that seem good to one will be the things that really are good for one. In this explanation, goodness must hold the logically prior position if it is to be the case that “the sound man is the sort of man for whom things absolutely good are good, on account of his own virtue” (*Politics* 1332a23). It will be with reference to the good that a person is determined good, as Aristotle claims in this near-definition: “a man is good for whom the things good by nature are good” (*EE* 1248b25).

### 21.4.2 *The haplôs boulêton*

What makes something truly and absolutely good also absolutely desirable? The argument against View (P) consisted primarily of the charge that it precludes the existence of a “natural” object of wish. It is part of Aristotle’s own view that the real

good is the natural object of wish; the good for humans is connected to what is “by nature” good in a paragraph of the *Eudemian Ethics*:

...to [the good person,] the end, what is best without qualification, is good; what is against nature, on the other hand, and involves corruption is not the good, but the apparent good. . . . Similarly, too, wish is naturally of the good, but also, against nature, of the bad, and one naturally wishes for the good, but, against nature, and through corruption, also the bad. (*EE* 1227a18–31)

The terms “nature” and “corruption” bring to mind Aristotle’s (and Plato’s) analogy comparing the health of the human body to virtue of the soul, and provide insight into another important section of our main passage. At *NE* III.4 1113a27–30, Aristotle illustrates the difference between good and bad persons using this analogy (I place the statement of his own account of wish, which precedes the analogy, in brackets):

[But if, then, we are not content with these views, should we say that the good is without qualification (*haplôs*) and in truth (*kat’ alêthian*) the object of wish, whereas what appears good to a given person is the object of wish for that person?]

We shall then be saying that for the person of excellence the object of wish is the one that is truly so, whereas for the bad person it is as chance will have it, just as on the physical level too the things that are truly healthful are healthful for people in good condition, whereas a different set of things is healthful for those that are diseased. . . .<sup>9</sup>

Aristotle’s reference to the truth brings us to a related issue that complicates our initial question about the object of wish. Is the good according to Aristotle a good to which the good person has cognitive access, or is the human good defined and determined in his system by what is good for the good person? Can Aristotle meaningfully distinguish the real good from the apparent good? Aristotle more than once speaks of the *spoudaios* as the *measure* of what is good or desirable or pleasant. In our key passage, he states that, “what most distinguishes the good person is his ability to see what is true in every set of circumstances, being like a carpenter’s rule or measure for them” (*NE* 1113a33–34). One may suppose that this remark stems from a view of goodness according to which the good is essentially bound up with what the good person *takes to be good*, and even indistinguishable from it.

Moreover, Aristotle’s description of the good person as a measure of what is good recalls the doctrine of Protagoras, famously rejected by Socrates, that “man is the measure of all things.”<sup>10</sup> Aristotle seems to be arguing that the good is what the “good” person finds good, the good person being the reference point for other determinations of goodness. The analogy to health is especially suggestive of this view: different things *are* healthy for healthy bodies than are for diseased ones. What is “truly” healthy will be what is healthy for healthy bodies. The way the analogy is taken results in vastly different readings of the larger passage: if the “truth” about health or happiness is whatever the body or mind in good condition declares it to be, conceptions will form the basis of the account in a fundamental way.<sup>11</sup> The determination of success in desire or belief will be based on the way that the successful person desires or believes. The good life for humans will be defined by the good person’s judgment about what is conducive to human flourishing. One could ask, then,

whether there is any substance to a directive exhorting us to know and seek what is truly good, except perhaps to suggest that we follow what the so-called “good person” does. Alternatively, Aristotle can be taken to hold that the truth of the matter is something the good person *has access to*. The good and the healthy are defined according to external constraints that also provide the criteria for identifying the good person. Irwin distinguishes the two ways of reading the reference to a measure as the ontological and the epistemological, the former being that “the good person’s approval constitutes something as good” (Irwin 1999, p. 207). Only the epistemological reading is plausible, however, given the multiple references to *the truth* that the good person *grasps*. Reading Aristotle’s reference to man as measure in just this way, Reeve finds it “hard to escape the conclusion that, for all the variations that there are in fine and pleasant things, there are yet unconditionally fine and pleasant ones of which the excellent person has unconditional knowledge” (Reeve 1992, p. 25). In what follows I will work through the remainder of Aristotle’s discussion of wish with the aim of defending this second interpretation of the idea of the good human as measure.

We can ascertain which argument Aristotle takes the health analogy to support by considering his remarks just prior to, and also following, the analogy. When Aristotle says that “for the person of excellence the object of wish is the one that is truly so, whereas for the bad person it is as chance will have it,” he suggests a distinction between what is truly *boulêton* and what is *boulêton* to each (*NE* 1113a26, quoted above). However, the Greek text leaves ambiguous the reference of “is . . . so.” Is Aristotle saying that the good person wishes for the truly good, or for the true object of wish? Irwin understands the latter, and reads the passage this way: “For the excellent person, then, what is wished will be what is [wished] in reality” (*NE* 1113a25). If we supply “wished” (*boulêton*), also following Ross and Broadie and Rowe, a straightforward distinction results: the good person wishes for the true object of wish, while the bad person wishes for any chance thing.

Aristotle is certainly drawing a contrast between good and bad persons, but there is a remaining uncertainty as to what kind. To reconsider the analogy to health: “just as with bodies, for those in healthy condition the truly healthy things are healthy, while other things are so for diseased ones,” Aristotle claims (*NE* 1113a26–27). But this analogy may seem to obscure our understanding of Aristotle’s distinction between the virtuous and the vicious. If the analogy is to bodies for which different things are healthy, Aristotle must be taken to be saying that different things are good *for* (not *to*) or desirable *for* different people.<sup>12</sup> Either is incompatible with Aristotle’s claims that there is a true object of wish, that the good is this object, and that it is therefore good and desirable for everyone.

We must not fail to notice that Aristotle sums up these examples with the general claim that, “concerning all things the good man judges rightly,” meaning by this that, “in each area the truth appears to him” (*NE* 1113a30). Thus the context in which the health analogy is given suggests that it is not Aristotle’s aim to comment on the dietary requirements of the sick person. He wants to distinguish not what *is* good for good as opposed to bad persons, but what *appears* good to each type of person. In saying that the good person *judges* rightly, Aristotle brings out the

relevant contrast between good and bad persons: to the bad person it is not the true good that appears good, but any chance thing. Thus to arrive at the basis for Aristotle's stated distinction between what is *boulêton* to each type of person, we can supply "apparent" (*phainomenon*) in the crucial sentence. In doing this we supply a premise that explicitly connects the *phainomenon* with the *boulêton*, and matches up with the main point of the health analogy. The result is that Aristotle reasons as follows: for each person, the apparently good is wished. To the good person, the apparently good is the truly wishable (which just *is* the truly good), while for the bad person anything might appear good. It then immediately follows, as Ross and Irwin note, that what is wished by the good person is the true object of wish, while what is wished by the base person is any chance thing, for these are the things that appear good to each type of person. The comparison will be to the healthy person who takes truly healthy things to be healthy, as opposed to the sick person, for whom unhealthy things seem healthy (see note 13).

It is not uncommon for Aristotle to search for a "true" object of a faculty. Deliberation, which also follows reason, will not be about just any object; we ought not call *bouleuton* what the "idiot or madman" deliberates about, he claims, but that about which "the sane person" deliberates (*NE* 1112a20–21). Yet we surely *do* call even what the foolish or mad deliberate about objects of deliberation; otherwise it would make no sense to endorse other objects of deliberation. The similarity to the discussion of *boulêsis* cannot be overlooked. In particular, the claim that the true *boulêton* is not the apparent good comes to mind; the apparent good is clearly *an* object of wish, on Aristotle's own admission, and even, in a way, the object of every wish. So it must be something else that Aristotle has in mind when he restricts the objects of *boulêsis* or *bouleusis*. Only one possibility seems to me a good one; the *proper* object of *bouleusis* will not be what the foolish deliberate about, nor will the *proper* object of *boulêsis* be any apparent good (i.e., what the fool wishes for). What Aristotle calls the "true" subject of deliberation will be whatever is most suited to the reasoning capacity, and because this capacity is most developed in the wise person, this person is a good indicator of this object.

To a sick body, water might seem healthy while it is in fact harmful; one may show all bodily signs of needing water, so that water would seem to be just the thing to relieve dehydration (and not merely to bring satisfaction of thirst), even if in reality it is not. So the relation expressed by the analogy is between the person of sound constitution and the object that this person finds, or judges, wholesome. For health, "things advantageous to a healthy body we pronounce good for the body absolutely," and for virtue: things good for the virtuous person are the truly good things (*EE* 1235b34). "*Haplôs*," as we see here, can plausibly be taken as "absolutely," "in fact," or "in truth."

We have so far ignored the other examples Aristotle gives to supplement the analogy to health; he also draws an analogy to sensible features like sweetness. Which reading do these cases fit—the one that has Aristotle identifying the object that *is* good or healthy, or the one that takes him to be noting what *seems* good or healthy? The cases of sweet/hot/heavy present a problem in that (at least some of) these are relative to a given person; what sense is there in calling a thing truly heavy,

for example? If something tastes sweet to a sick person, in what way exactly could one maintain that it is in fact bitter? In fact, it is hard to see how the healthy and the sweet are supposed to be analogous at all. While it is not so problematic to suppose that there is something truly good or noble and that the wise and prudent person can ascertain what this is, for the pleasant it is not so simple, and for the sweet or heavy it is even less so. Nevertheless, Aristotle maintains that there is a truth of the matter in each of these cases.

The analogy is most effective if we consider that Aristotle held that there is some kind of judgment involved even in the case of tastes. As he explains in the *De Anima*, “each sense. . . judges the varieties of the subject perceived by it, e.g., sight for white and black, and taste for sweet and bitter; and similarly for the other senses too” (426b10). It is also the sensory faculty (*aisthêsis*) that can judge (again, *krinei*) and discern differences in objects that have some quality. There is a truth about what is healthy, bitter, sweet, etc., that parallels the case of virtue, in which the good person’s wish corresponds to the truth (also the ideal or standard) for goodness. The healthy person will find really bitter things to be bitter, really sweet things to be sweet, and really hot things to be hot, as he “sees the truth in each kind” (*NE* 1113a25). The appeal to well-functioning is unmistakable in Aristotle’s claim that “the same thing never appears sweet to some people and the contrary of this to others, unless in the one case the sense organ which judges the said flavors is injured or defective” (*Met* 1063a2–4). The truly healthy is healthy for everyone, as is the truly desired desirable for all, though not desired by all, and the truly sweet seems sweet to the healthy person but not necessarily to a diseased one—to the diseased, bitter things seem sweet but aren’t. Similarly, judgment and wish in the virtuous person function as they should ideally and by nature, and as if defective or diseased in the vicious person. Thus, Aristotle can reasonably suggest that the healthy person registers flavors as they are. Lemons are sour to the well-functioning person in ordinary circumstances. If lemons are sweet and honey is bitter to you, they are sweet and bitter *to you*, but not *really*. Indeed, these responses to these foods may cause you to realize that you are not apprehending things as they are.

Aristotle’s conclusion reveals his aim in constructing the analogy: “the good man judges everything correctly; what things truly are, that they seem to him to be, in every department” (*NE* 1113a25). Moral virtue is one area among many, within which human faculties function correctly by attaching to the object that correct judgment has identified to be truly or absolutely of the kind required. This means that when the faculty of wish is functioning correctly, and optimally, it takes as its object the truly excellent. Failure in judgment or cognition of the truth results in the pursuit of something that seems good to a given person but is not absolutely good—it is only an apparent good. Just as in the case of a diseased body, to which a sweet thing seems bitter, to a vicious person a bad end seems good.

The distinguishing mark of the good person is therefore that “the truth appears to him” (*EE* 1248b25). Aristotle’s doctrine of the good person as a measure of good things is a fitting way to capture the relationships between wishing and the good. Because of her access to the truth, the good person is a standard and measure of the good and the pleasant in every case (*NE* 1113a32).



It would seem that if Aristotle derived goodness in the human sphere from the virtuous person's assessment of and reaction to the world, this would be most clear in the case of pleasure. That is, it seems reasonable that if any pleasures can be called best among pleasures, these will be whichever ones the good person finds pleasant; thus the good person will be in the full sense a standard, and not just a measure, of the pleasant. It would be the case that were pleasures simply weighed against each other, no differences in kind would present themselves; pleasure would be pleasure whether complementing study or intoxication.<sup>13</sup> Correspondingly, better activities would not bring better pleasures, though differences in degree of pleasure would be felt within and across activities.

But we find for pleasure, as for wish, that there is something "by nature" pleasant, according to Aristotle. Again we see that this may mean one of several things: (1) that the "truly" pleasant is determined by what the virtuous person finds pleasant, (2) that the virtuous person, unlike others, finds the truly pleasant to be pleasant, and so gets it right, or (3) that there's no value attached to pleasure; pleasure attends good or bad objects, but admits of no corresponding difference in the pleasure itself. Appropriateness is found in one's taking pleasure in certain activities and not others.

I think Aristotle maintains the second position. It would be strange to speak of the "truly" pleasant if the pleasant were not so determined. What is added to the requirement that a person find pleasure in the right things by maintaining nonetheless that the pleasure that attends them is no different in quality but is "better" simply because the good person feels it? The account requires that the pleasures of virtue are qualitatively better than those of the body, and specifically so *to a human*, a creature most suited to activity in accordance with the rational element.<sup>14</sup> This is the explicit point of Aristotle's adding that virtuous actions are pleasant to virtuous people, "as well as *in their own nature*" and again that "virtuous actions must be *in themselves* pleasant" (*NE* 1099a15–20).

The account of pleasure supplements the account of desire in a more explicit way: we find that the good person "judges well" about the goodness, fineness, and pleasantness of good actions (*NE* 1099a23: *kalôs krinei o spoudaios*). Aristotle's treatment of the inscription at Delos suggests that he thinks of pleasure as a quality of activities, as something they contain: the best activities *possess* fineness, goodness, and pleasantness. We return to a distinction of Socrates' in Aristotle's claim that the life of virtue is "in itself" (*autên kath' autên*) pleasant, which must be equivalent to its having pleasure in itself (*NE* 1099a16).

As in the case of wish, Aristotle maintains the existence of the pleasant by nature, against the unacceptable alternative that contrary objects can have the same status with regard to the pleasant. We have in the account of pleasure a parallel to the obscure characterization of wish that sees as the alternative to a "true" and "natural" object of the faculty the situation in which "even contrary things" are objects of wish. The contrary to an object of wish cannot be equally good as an object of wish. So too for pleasure: what happens in the case of most people is that their pleasures are in conflict. This is *because* the pleasures of most are not pleasant by nature, whereas to the lover of virtue the things that are pleasant are the truly pleasant things.

Among truly pleasant activities are virtuous acts, so that these are “pleasant *both* for virtue-loving people *and* in themselves.” Aristotle is clearly making *two distinct points* here: being truly pleasant, virtuous actions are pleasant strictly and relatively; they are themselves pleasant and also pleasant to those who undertake them. This is why, as for other matters, the good person is a measure of what is truly pleasant, while what she finds pleasant is one measure of her goodness (*NE* 1099a21). The judgment of the good person is again an *indicator*: Aristotle claims that correct judgment will produce virtuous actions that are most fine as well as most pleasant. He takes it to have been established that the good person *does* judge correctly (*kalôs krinei*), which means that the good person’s judgment coincides with the truth of the matter (*NE* 1099a22). Aristotle’s overall point is clear enough: human goodness brings with it the greatest and most fitting pleasure for humans; this is revealed in those who are virtuous.

I have presented Aristotle’s characterization of the “natural” object of wish within the framework of the connections present in the abstract, without considering evidence of an appeal to the nature of the human being. It would provide a basis for the account of natural objects of a faculty to understand the fulfillment of a human or its activities as involving the attainment of this object. The obvious place to look for this foundation is in Aristotle’s function argument, which is based on the assumption that the good for humans will be found in the well-functioning of humans.<sup>15</sup> The function Aristotle seeks is a characteristic activity to which humans are naturally suited. It will, in addition, be some part of a human’s nature that is particular to the kind of creature it is: activities reflecting the exercise of the reasoning capacity. Goodness for a human is consequent on the successful exercise of the faculty, which in turn depends on the quality that allows humans to function well. In this consists perfection and well-being.

When Aristotle examines human activities to determine which one of them is the peculiar human function, he appeals exclusively to natural capacities.<sup>16</sup> Santas points out what is for our present purposes the key feature of the concept of function: that it “makes essential reference to characteristic capacities or potentialities of an object and the activities to which these issue, and not necessarily to human desires or ends” (Santas 1989, p. 89). In fact, there is *no* appeal to desire in this part of the *Nicomachean Ethics*, but Aristotle’s account of desire complements his account of function, as desire is the faculty by which we are attracted to just those things that promote the essentially human activity of the soul.

A related argument connects the different lines of thought in Aristotle: the conception of vice as destroying the soul in the way that disease destroys the body indicates that it is the nature of the soul that forms the foundation of Aristotle’s insistence on the necessity of virtue for well-being (*NE* 1104a10). Aristotle determines for the body and the soul what each is, what it is suited to do, and what allows it to do this well or causes it to do it badly.<sup>17</sup> The goal is to narrow down which excellence is particularly human, but Aristotle avoids characterizing the possession of virtue, fully realized, as natural: because we are “adapted by nature to receive” the virtues, they are neither fully natural nor against nature (*NE* 1103a25). The nature of the human soul and the sort of well-being to which it is suited are

unique: no other animal can be happy because none can participate in this kind of activity, the activity of soul in accordance with virtue, as has just been specified (*NE* 1099b31). This is evidence as to the basis of Aristotle's theory: how it is determined what happiness we are talking about, and what goodness.

### ***21.4.3 Aristotle's Position on the Idea that "All Desire is for the Apparent Good."***

The role played by the apparent good in the account is, as we have seen, to explain how all cases of wishing are exercises of the faculty. But the apparent good is also a key concept in the explanation of the case of *error*: "in contravention of nature and by perversion not the good but the apparent good is the end" (*EE* 1227a23). Aristotle's explanation of how this can be reveals a specific use of the distinction between real and apparent. The apparent good is here thought of as the *merely* apparently good, *not* simply what appears good, which could just as well *be* the real good:

The reason is that there are some things that cannot be employed for something other than their natural objects, for instance sight—it is not possible to see a thing that is not visible, or to hear a thing that is not audible, but a science does enable us to do a thing that is not the object of the science (*EE* 1227a24–26).

With this clarification we return to the two sub-questions about wish that framed Aristotle's inquiry: "What is the actual object of wish?" and "What is the ideal object of wish?" Wish is given immediately following the quoted passage as an example: one can exercise the faculty without exercising it toward the right object (its natural object). Thus by "apparent good" Aristotle here means apparently but not really good, i.e., bad.

Evans illustrates the case of wish as follows: "Shooting and targets are essentially correlative; without the one the other cannot exist. But only some shots hit at the target; and of those which have failed to hit we may say that while they *were intended* to be directed at *the* target, they were in fact directed at what they hit, which was therefore *their* target" (Evans 1987, p. 59). The distinction looks very much like one that is useful in exploring interpretations of Plato's Socrates—namely, that between an *intended* and an *actual* object of wish, as explained by Santas.<sup>18</sup> But it also functions at another level: we *want* for all of our desires to be directed at the real good, but often they aren't, so this higher aim is also frustrated. Charles explains that "the practically wise fulfill the goal of practical reasoning because they do what is best (non-accidentally) as they see what is best, desire it appropriately, and desire no other action more than it" (Charles 1984, p. 189). In contrast, those with other states of character will make mistakes in identifying the good, or alternatively happen upon it but without the relevant knowledge.

The Socratic view of the *Gorgias* presents a stark contrast to Aristotle's view on one of the questions that frame the inquiry in the *Nicomachean Ethics*: the actual object of desire is what appears good to each person. While it is true on Aristotle's

account that, descriptively, everyone desires what appears good, this is not at the same time an adequate account of the faculty: it is incomplete as a characterization of wish (conceived in the abstract) or the desires of the virtuous person. Aristotle completes the account by relating wish to the real good, and emphasizing the knowledge and responsibility that are involved in the virtuous person's desires and actions. Aristotle does not characterize the predicaments of the virtuous and non-virtuous as equally mired in obscurity; the wise are not fumbling about hoping for success, but attuned to the reality of that which they seek. Though Aristotle characterizes the ends that ignorant people wish for as merely apparently good, such people do not differ from the wise with respect to their ultimate aim—i.e., happiness. Though they may be vastly mistaken as to the nature of happiness, it would not be correct to suppose that their ultimate desire would be satisfied by the acquisition of whatever they happen to identify with happiness. It appears that, for Aristotle as well as for Socrates, we all desire to be happy even if we do not know what constitutes happiness. Thus the two views do coincide on at least two points: rational desire aims at the real good, and knowledge is the determining factor in one's acquisition of that good.

Finally, I would like to elaborate upon Aristotle's distinction between what is in accordance with nature and what is against nature, by noting the role of belief in desire. For Aristotle and perhaps non-Aristotelians too, the concept of rational desire is bound up with that of goodness: to desire, in this way, is to seek to attain something *beneficial* (whereas to desire through appetite is just to crave). Because Aristotle believes that some beliefs, or "appearances" are better than others—specifically by being *true*—part of becoming virtuous is cultivating our abilities to discern correctly what is dangerous, healthy, and so on. Making explicit the relation desire bears to its true object by way of successful cognition, Reeve describes the person of practical wisdom as one who is able "to register things as they are and to care about them and be moved by them proportionately to their true value and importance" (Reeve 1992, p. 196). I understand Aristotle's view to be that goodness, broadly construed, aligns us with reality in several ways: we apprehend reality correctly, desire things that are actually good, excel in characteristically human ways, and experience happiness of a kind and to a degree that particularly suits our human nature.

**Acknowledgements** It is truly a pleasure to contribute to the present volume in honor of Jerry Santas, whose dedication to the study of Plato and Aristotle has shaped my deep appreciation for the insights of the ancient Greek thinkers. I am fortunate to have, in Jerry, a mentor who has so generously encouraged and challenged my thought over so many years. I wish to thank both Jerry Santas and Georgios Anagnostopoulos for helpful critical comments on the present paper.

## Notes

1. Throughout the argument with Polus at *Gorgias* 466b–468e, Socrates insists on the distinction between doing what one wishes and doing what one thinks best, indicating that one does not necessarily desire what appears good to one. This exchange opens the argument: "POL. Then they do not do what they wish? SOC. I say no. POL. When they do what they think fit?"

- SOC. Yes” (467b3–9). And the argument ends with Polus accepting that someone who does something, “because he thinks it better for himself though it is really worse. . . does what he thinks fit,” but does not do what he wishes (468d1–10).
2. Because the virtuous person apprehends and desires the real good for what it is, I am inclined to think that what “appears” to this person, in Aristotle’s view, is just the good, and not the good under the description “good.” However, I am not able to explore here the implications of Aristotle’s commitment to the existence of conceptions for the possibility that knowledge of the good relates the wise person to reality directly, rather than by way of such conceptions. I am most grateful to Terry Penner for illuminating the complexity and significance of this issue, and hope to address it elsewhere in the future.
  3. *NE* 1113a19. In the *Gorgias*, Socrates argues that we wish for the end, the good, unconditionally, and for particular actions conditionally upon their bringing us the good at which we aim.
  4. Thus, Aristotle would say that Polus was correct to question Socrates’ contention that one can think something is best for one but nevertheless *not* also want to do it.
  5. It is not open to Aristotle to say that the person who thinks his happiness is constituted by bodily pleasure aims for it because he has an appetite (*epithumia*) for pleasure. The desire for pleasure as an end must be a wish; appetite meanwhile will be for things that one does in order to attain pleasure (*NE* 1118b7–1119a5).
  6. Aristotle, *De Anima* 433a21. Aristotle states this same idea in various ways: at 432b15 he notes that, “nothing that is not desiring or avoiding something moves unless as a result of force,” and at 433b10 he concludes, “hence that which produces movement will be one in kind, the faculty of desire as such—and first of all the object of desire.”
  7. Penner and Rowe (1994, pp. 4–5) take this passage to show that, for Aristotle, every case of desiring involves perceiving an object under the description “good,” and to bring out the difference between Socrates and Aristotle: for Socrates desire is not mediated by appearances of objects, while for Aristotle it is, in every case, even when the apparent good is the real good.
  8. Irwin (1999, p. 318), note on *NE* 1113a15ff. This can be seen in the varying translations of Aristotle’s formulation of view (P): “the apparent good is wished” (Irwin 1999), “what appears good is wished for” (Rackham 1926), “the apparent good is the object of wish” (Ross 1984).
  9. The Greek reads: *tois men eu diakeimenois hugieina esti ta kat’ alêtheian toiauta ontâ*. . . and is thus ambiguous between two ideas: that truly healthy things are healthy *for* healthy persons, and that truly healthy things are *taken to be* healthy *by* healthy persons. Rackham, for example (retaining the ambiguity), translates the passage in this way: “a man of sound constitution finds really healthy food best for his health, but some other diet may be healthy for one who is delicate.” I address this below.
  10. The view is discussed in Plato’s *Theaetetus* 151 ff. and Aristotle’s *Metaphysics* Δ.4.
  11. This would mean that when Aristotle says “the real good” is desired, he can only mean what appears good to the good person, and therefore endorses an altogether different distinction between the real and the apparent good.
  12. Irwin (1999, p. 207) notes that, “it is not clear whether Aristotle is saying (i) Broccoli is really healthy for healthy people, and antibiotics are really unhealthy for them. . . or (ii) Healthy people judge correctly that broccoli is healthy for them. . . .” I am here agreeing with Irwin that “the sensory examples that follow, and the demands of the argument as a whole, suggest that here Aristotle has (ii) in mind.”
  13. Plato presents this question in the *Philebus* asking, “whether it is at all possible that, as against other things that have quality, pleasure and pain never have qualities but simply are what they are.” Socrates concludes that pleasures can be false, for example by corresponding to false opinion. Furthermore, “bad men delight for the most part in false pleasures, good men in true ones” (36e and 40c; Hackforth transl.).
  14. Plato gives an argument for this in the *Republic* based on the different pleasures peculiar to the different parts of the soul: the pleasures of Reason are the truest and the most pleasurable (580d–587a).

15. Keyt (1983, p. 367) expands upon Aristotle's explicit claim that the good for a thing will be found in its function to reach a premise that is needed in the argument: "the good for—that is to say, the ultimate end of—a member of a kind is to be a good member of its kind."
16. A similar view can be seen in Plato's function argument (the function of the human soul is to deliberate and rule, and only by doing these well can it be happy) and in the way he defines Justice: it requires that everyone "pursue one occupation of those in the city, that for which his nature best fitted him" (*Republic* 353d–e and 433a).
17. *De Anima* 415b16: Nature operates always toward some end, and the soul supplies the end in accordance with nature for living things. Also, the body is an instrument of the soul, and its purpose is the benefit of the soul. This is to make clear the foundation of Aristotle's theory of soul, and to suggest that both its functioning and its well-functioning are rooted in the conception of what a human being is.
18. What Santas (1979) calls the "actual" object of desire is identifiable according to the true qualities of the object, and independently of the way one construes it. The "intended" object of desire, on the other hand, is identified by way of the description under which the object is sought.

## Bibliography

- Broadie, S., and C.J. Rowe. 2002. *Aristotle: Nicomachean Ethics*. Oxford: Oxford University Press.
- Charles, D. 1984. *Aristotle's Philosophy of Action*. New York: Cornell University Press.
- Evans, J.D.G. 1987. *Aristotle*. New York: St. Martin's Press.
- Forster, E.S., trans. 1960. *Aristotle: Topics*. Cambridge: Harvard University Press.
- Grant, A. 1973. *The Ethics of Aristotle*. New York: Arno Press.
- Hackforth, R., trans. 1971. "Plato: Philebus". In *Plato: The Collected Dialogues*, edited by E. Hamilton. Princeton: Princeton University Press.
- Hamlyn, W.D., trans. 1968. *Aristotle: De Anima*. Oxford: Clarendon Press.
- Irwin, T., trans. 1979. *Plato: Gorgias*. Oxford: Clarendon Press.
- Irwin, T., trans. 1999. *Aristotle: Nicomachean Ethics*. Indianapolis: Hackett.
- Joachim, H.H., trans. 1962. *Aristotle: Nicomachean Ethics*. Oxford: Clarendon Press.
- Jowett, B. 1968. "Aristotle: Politics". In *The Basic Works of Aristotle*, edited by R. McKeon. New York: Random House.
- Keyt, D. 1983. "Intellectualism in Aristotle." In *Essays in Ancient Greek Philosophy*, edited by J. Anton and A. Preus. Albany: State University of New York Press.
- Penner, T., and C.J. Rowe. 1994. "The Desire for Good: Is the *Meno* Inconsistent with the *Gorgias*?" *Phronesis* XXXIX:1.
- Rackham, H., trans. 1926. *Aristotle: Nicomachean Ethics*. Cambridge: Harvard University Press.
- Reeve, C.D.C. 1992. *Practices of Reason: Aristotle's Nicomachean Ethics*. Oxford: Clarendon Press.
- Roberts, R., trans. 1968. "Aristotle: Rhetoric". In *The Basic Works of Aristotle*, edited by R. McKeon. New York: Random House.
- Ross, W.D., trans. 1984. "Aristotle: Nicomachean Ethics". In *The Complete Works of Aristotle*, edited by J. Barnes. Princeton: Princeton University Press.
- Santas, G. 1979. *Socrates*. London: Routledge & Kegan.
- Santas, G. 1989. "Desire and Perfection in Aristotle's Theory of the Good." *Apeiron* 22(2):75–99.
- Shorey, P., trans. 1937. *Plato: Republic*. Cambridge: Harvard University Press.
- Tredennick, H., trans. 1962. *Aristotle: Metaphysics*. Cambridge: Harvard University Press.
- Woods, M., trans. 1992. *Aristotle: Eudemian Ethics*. Oxford: Clarendon Press.

## Chapter 22

# Continuity and Incommensurability in Ancient Greek Philosophy and Mathematics

Vassilis Karasmanis

Aristotle, in the sixth Book of his *Physics*, discusses extensively the problem of continuity. Continuity is the main and essential characteristic of magnitudes. The continuum is found only in magnitudes and all magnitudes are continuous. Both philosophers and mathematicians distinguish between arithmetic and geometry in the sense that arithmetic deals with numbers and geometry with magnitudes.<sup>1</sup> Magnitudes are continuous; numbers are discrete. In the *Categories* (4b20–24) Aristotle says that “of quantities some are discrete (*diôrismenon*), others continuous (*suneches*). . . . Discrete are numbers and language; continuous are lines, surfaces, bodies, and also, besides these, time and place.”<sup>2</sup>

For Aristotle, the essential characteristic of continuity is infinite divisibility.<sup>3</sup> If a line is continuous, then we are able to cut it at any point we want. If we are able to cut a line at any point, then we have infinite divisibility, because between any two points A and B there is always another point C. If a line is not infinitely divisible, it is not continuous. For Aristotle, the exemplary case of infinite divisibility is the Zenonian dichotomy.<sup>4</sup> However, another significant characteristic of magnitudes is incommensurability. Because of that, incommensurability appears only in geometry and not in arithmetic. Aristotle, in his treatment of the continuum, does not refer to incommensurability at all, although he knows the phenomenon very well.<sup>5</sup> Why does this happen? Does he think that incommensurability is a result of infinite divisibility? We do not have evidence that Aristotle thought so. On the contrary, from a passage in the *Prior Analytics* (65b16–20) we see that Aristotle thinks that infinite divisibility and incommensurability are two rather independent properties. In this passage, he refers to a fallacy according to which we “posit as a cause that which is no cause;” and he gives the following example:

[A]s if someone wishing to prove that the diagonal of a square is incommensurable were to try to prove Zeno’s argument that motion is impossible, and were to use reduction *ad impossibile* to this end; for there is no connection in any way at all between the fallacy and the original assertion.

---

V. Karasmanis (✉)  
National Technical University of Athens, Athens, Greece  
e-mail: vasmar@hol.gr



Now, the original assertion is about the incommensurability of the diagonal and the fallacy is Zeno's argument against motion, which relied on infinite divisibility, and Aristotle says that there is no connection between these two.

So Aristotle's continuum, which relies on Zenonian infinite divisibility, does not exhaust magnitudes. A Zenonian infinite divisibility never cuts a line into incommensurable segments. However, we do not have evidence that Aristotle had understood this problem.<sup>6</sup> On the contrary, it seems that he thought his "continuity" (and therefore Zenonian infinite divisibility) and of "magnitudes" as equivalent terms. His only distinction is between continuous and discrete; and that magnitudes are continuous and numbers discrete (see e.g., *Physics* 200b 17–25, 231b 19–23; *Categories* 5a 1–7). Aristotle's approach to the continuum is rather physical (see also Düring (1996), p. 62). It starts from his effort to support the simple fact in the physical world that there is motion and change. Nevertheless, is another approach to continuity and magnitudes that exploits the tools of ancient mathematics possible? In other words, is it possible to attempt to understand and analyze these notions starting from the phenomenon of incommensurability? I shall try to offer arguments in support of the claim that Plato's treatment of *apeiron* (indeterminate or unlimited) in the *Philebus* shows such an effort.

Plato, in the *Philebus*, speaks about the terms *peras* (limited or determined) and *apeiron*<sup>7</sup> in two main passages: 16b–19a and 23c–27c. In the first passage Plato speaks of a method as a "gift of the gods" (16c5)<sup>8</sup> and continues by saying (16c9–10) that "the things that are ever said to be (*tôn aei legomenôn einai*) are made up of one and many, with a determinant (*peras*) and indeterminacy (*apeiron*) inherent (*symphyton*) in them." In 23ff he makes a fourfold division *peras*, *apeiron*, mixed (*meikton*) and cause (*aitia*), saying that "the god, we were saying, I suppose, showed one [part] of the things that are as unlimited, another as limit" (23c9–10); and he goes on to examine these four categories one by one. These two passages have created many problems of interpretation. (1) What are the "things that are ever said to be" in 16c9 and the "things that are" in 23c9? (2) Are *peras* and *apeiron* the constitutive principles of every thing (16c9) or two general classes of things (23c9–10)? (3) Does the *apeiron* in 16b–19a refer to the unlimited particular instances of a Form or to unlimited types of a Form? Is also the *apeiron* in 16b–19a the same with the *apeiron* in 23c–27c? (4) What is the role of the Platonic Forms in these two passages? (5) Does the passage of the fourfold division refer to Forms, to the sensible world or to both? (6) If this passage refers to Forms, in which of the four categories do they belong? I think that we shall be able to answer the above problems only after a systematic examination of the characteristics of *peras* and *apeiron*. In this paper I shall restrict myself only to the examination of Plato's *apeiron* and its characteristics. I shall refer mainly to passage 23c–25b where Plato speaks of *apeiron* and *peras* independently but also to some later passages where Plato speaks about *apeiron* in the framework of his discussion on the mixed (*meikton*). I am not going to refer to the first passage (16b–19a) because the information we have about *apeiron* there is very scarce and any attempt to say something about it necessarily involves an interpretation of the whole passage.<sup>9</sup>

At 24a–b, Plato says regarding the *apeiron*: “See if you can discern any determinant (*peras*) in relation to hotter and colder; or is it true that so long as the more and less that inhabits these categories (*genê*) continue there they would not permit any end to come about? Indeed any such ending would be the end of them? —True enough. —Now, we agree that hotter and colder always contain more and less? —Yes. —The argument suggests that these two are always without end. Being without end they are completely indeterminate (*apeira*.)” And, at 31a he says that the *apeiron* “in itself does not have and will never have any precisely marked beginning, middle, or end.”

So far, *apeiron* seems to be something that is “more and less” and does not have a limit in the sense of a beginning, end, middle or any other distinguishing point. In this way, for the pair hotter/colder, for any hot or cold we have always something hotter or colder. Also, Plato does not speak of hot and cold but of hotter and colder; and that the hotter and colder contains always the “more and less”—not the “more or less.” This means that any hotter is more and less and in this way it is also colder. They are all comparatives. The use of the comparatives means if we have two points on a line we can always compare them and find which is hotter and which is colder. The two directions of the line are determined: the direction of the hot and the direction of the cold. Also, there is no demarcation between cold and hot. We can therefore represent this *apeiron* (hotter-colder) as an infinite line open to both directions. Also, in this line there is no point—no zero point, as it were—that demarcates the hot from the cold.<sup>10</sup> However, from the above passages, we cannot conclude whether this line is continuous or discontinuous.

But at 24c–d, Plato explains better what he means by “*apeiron*” and “more and less.” He says that wherever the more and less “are present they exclude any definite quantity (*poson*)” and

[I]f they do not obliterate any definite quantity but allow degree and measure (*metron*) to appear in the midst of more and less, they in fact abandon the territory they occupied. For in admitting of definite quantity, they would no longer strictly be hotter or colder. For the hotter goes on without pause, and the colder in the same way, while a definite quantity comes to a particular point and goes no further. So on the present argument the hotter and at the same time its opposite would come out as indeterminate (*apeiron*).

This passage says that the essential characteristic of the “more and less”<sup>11</sup> is the absence of any definite quantity, for the presence of definite quantity and measure in the place where the more and the less is present will destroy the more and the less. It seems that according to Plato the notions “more and less” and definite quantity are mutually excluding. Now, something that does not permit definite quantities also excludes measure and number; moreover such a thing cannot consist of discrete particles or atoms. For, if our line consisted of discrete atoms, we would have a natural measure on it. What is characterized by the more and less should be continuous and, therefore, the characteristics that Plato attributes to the *apeiron* point to continuous magnitudes. In addition, the expression “goes on without pause” denotes a continuous motion. Such an *apeiron* as continuous is infinitely divisible, and this last property is the main characteristic of magnitudes. We see the same idea in what he goes on to say (24e–25a) right after the passage just quoted.<sup>12</sup>

However, there is a passage that complicates our understanding of Plato's *apeiron* as continuous. At 27e, speaking about pleasure, Plato says that it is *apeiron* both in quantity and degree (*kai plêthei kai tô(i) mallon*). The word *plêthos* denotes the sum of discrete things. For example, in Euclid's *Elements* number is defined as "sum [*plêthos*] of units." In this way *apeiron* in *plêthos* is something that is infinite by addition (following Aristotle's terminology; see *Physics* 204a6–7), like the series of numbers. This *apeiron* is not continuous but discrete. On the other hand, the *apeiron* in degree (*tô(i) mallon*) does not have definite quantities and, therefore, it is continuous or infinite by division.

Let us now summarize the characteristics of Plato's *apeiron* as accepting the "more and less": (1) It is something that does not have any specific beginning or end. In this way, it can be represented by a line open in both directions. (2) These two directions are determined by the two opposites of the *apeiron* (hotter-colder). (3) There is no definite quantity or measure in it. From this we can conclude that "for any measure we may take there is always a smaller one," and also that "for any two points in such a line there is always a point between them." This suggests the fourth characteristic which is (4) infinite divisibility and therefore continuity. (5) There is no point zero in the line of *apeiron*. There is no fixed demarcation between, e.g., hot and cold (for some of these characteristics, see Meinwald 1998, p. 173).

Nevertheless, these characteristics do not exhaust Plato's conception of *apeiron*.<sup>13</sup> There is another, very relevant passage where Plato speaks about *peras* in opposition to *apeiron*:

Things that do not allow these features, but admit of all the opposite things—equal and equality, and after that double and everything that is number to number or measure to measure, all these we should be advised to apportion to *peras*. (25a6–b2)

Let us make some comments on the above passage: "Equality" and "double" are relations between two numbers or other things. These relations can be expressed by arithmetical ratios. Therefore, the expressions "number to number" and "measure to measure" are generalizations denoting not only these two relations but all arithmetical ratios. In addition, Plato does not say only "number to number" but he adds "measure to measure." The formula "number to number" applies only to things that are discrete by their nature, e.g., numbers, oranges, and oxen. But in everyday life we treat continuous magnitudes as discrete things by imposing on them an artificial measure. So we count distances by feet or meters, weights by kilograms, time by hours or minutes, etc. It seems then that Plato adds the formula "measure to measure" in order to cover all these cases of everyday calculations or practical mathematics.

I come now to a more serious point. Plato's description—or definition—of *peras* as "number to number" is exactly the same as what mathematicians understand by commensurable magnitudes.<sup>14</sup> Let us look at some propositions from Book X of Euclid's *Elements*:

Proposition 5: "Commensurable magnitudes have to one another the ratio which a number has to a number."

Proposition 6: “If two magnitudes have to one another the ratio which a number has to a number, the magnitudes will be commensurable.”

Proposition 7: “Incommensurable magnitudes do not have to one another the ratio which a number has to a number.”

Proposition 8: “If two magnitudes do not have to one another the ratio which a number has to a number, the magnitudes will be incommensurable.”<sup>15</sup>

Since the theory of incommensurability we find in Book X of Euclid’s *Elements* was developed by Theaetetus within Plato’s Academy,<sup>16</sup> I find it very plausible that in the passage quoted earlier Plato has in mind the above very elementary propositions of the theory of incommensurability. So, it seems that Plato suggests that the main characteristic of his *peras* is commensurability. We find further evidence for this claim in two other passages from the *Philebus*: (a) “That of equal and double, and whatever puts an end to opposites being at odds with each other, and by the introduction of number that makes them commensurate (*summetra*) and harmonious” (25d11–e2). (b) “Again, in the case of extremes of cold and heat its advent removes what is far too much and *apeiron* and produces what is measured (*emmetron*) and commensurable (*summetron*)” (26a).<sup>17</sup> If *peras* is what makes things commensurate, then *apeiron* must be the source of incommensurability. I think that Plato is here using the term *symmetron* in a rather technical, mathematical sense.

Now, if we agree that Plato says (25a6–7) that *apeiron* admits opposite characteristics to those of *peras*, then we have to conclude that incommensurability is a further and very important characteristic of *apeiron*. It seems then that Plato relates discontinuity to commensurability and probably (*ex silentio*) continuity to incommensurability. Let us examine if this is true. If we define continuity in terms of Zenonian infinite divisibility, as Aristotle does, then all the pairs of the infinite cuts of a continuous thing (e.g., a line) have a specific ratio (a):(b) (where (a) and (b) are numbers). Therefore, according to Plato’s conception of *peras*, the Aristotelian continuum belongs to *peras* and not to *apeiron*. Hence, we cannot interpret Plato’s *apeiron* just in terms of continuity or infinite divisibility.<sup>18</sup>

It seems that Plato’s interest is not to identify his *apeiron* with the continuum. I think that Plato’s purpose is rather to examine and penetrate the notion of magnitude. The gift that the gods give to humans, at 16cff, is a method that enables them to create or discover arts or sciences (*technai*). The examples that Socrates gives at 17b–16d have to do with music and grammar. The notion of *technê* seems to play an important role in the *Philebus*.<sup>19</sup> The two main sciences at the time of Plato were arithmetic and geometry. Not only that but, as Plato admits, all other arts depend more or less on them (55e–56c). The early Pythagoreans believed that “everything is number”; but after the discovery of incommensurability, mathematicians realized that numbers and magnitudes are two essentially different things. In this way we have a sharp distinction between arithmetic, which has numbers as its subject matter, and geometry, which has as its subject matter magnitudes. *Peras*, according to Plato, is everything that can be expressed by numbers or arithmetical relations (25a7–b1). *Apeiron* is what does not accept numbers or arithmetical relations, but the “more

and less.” Magnitudes have all the characteristics Plato attributes to *apeiron*. When Plato says that all things are constituted by *peras* and *apeiron*, he probably has in mind everything that is discrete on the one hand, and continuous magnitudes on the other.

Plato’s approach to *apeiron* is rather mathematical. He approaches magnitudes via incommensurability and not via infinite divisibility. Of course, all magnitudes are infinitely divisible in the Zenonian sense, and therefore continuous. But magnitudes are also dense, in the sense that they include incommensurable cuts. Is there any relation between infinite divisibility and incommensurability? Let me continue my reasoning beyond the text of the *Philebus*.

If we ask how it is possible to find whether or not two magnitudes are incommensurable, the immediate answer, from the point of view of Greek mathematics, is “by *anthuphairesis*” (reciprocal subtraction). *Anthuphairesis* is an old mathematical technique<sup>20</sup> used mainly in arithmetic in order to find the greatest common divisor between two numbers. We read in Proposition X, 2 of Euclid’s *Elements*: “If, when the less of two unequal magnitudes is continually subtracted in turn from the greater, that which is left never measures the one before it, the magnitudes will be incommensurable.” Hence, an infinite *anthuphαιρεtic* process of reciprocal subtraction between two magnitudes shows that these magnitudes are incommensurable.<sup>21</sup>

In this way we see that incommensurability appears as special case of infinite divisibility. Let us call it *anthuphαιρεtic* infinite divisibility. We have, therefore, two kinds of infinite divisibility: (1) Zenonian infinite divisibility which results in continuity; and (2) *anthuphαιρεtic* infinite divisibility which produces the incommensurables and makes the continuum dense, thus generating magnitude. So supplementing the Aristotelian continuum—which is characterized by Zenonian infinite divisibility—with the incommensurables—which are found with the *anthuphαιρεtic* infinite divisibility—we get magnitudes.

Did Plato see all these results? I find it very improbable. But, since he had a good knowledge of the mathematics of his time and especially of the philosophical and foundational problems of it (see Karasmanis 1990), it is not unreasonable to suppose that he had an intuitive grasp of some of these problems and so preferred to approach magnitudes in terms of incommensurability rather than in terms of Zenonian infinite divisibility.

There is a passage in Proclus’ *Commentary of the First Book of Euclid’s Elements* which seems to advance similar ideas. Although Proclus does not refer to Plato at all, I find it highly probable that he has in mind our passage on *peras* and *apeiron* in the *Philebus*.

Mathematicals are the offspring of the Limit (*peratos*) and the Unlimited (*apeirian*), but not of the primary principles alone, nor of the hidden intelligible causes, but also of secondary principles that proceed from them. . . . This is why in these orders of being there are ratios (*logoi*) proceeding to infinity (*apeiron*), but controlled by the principle of the Limit (*peratos*). For number, beginning with unity, is capable of indefinite increase, yet any number you choose is finite; magnitudes (*megethôn*) likewise are divisible without end, yet the magnitudes distinguished from one another are all bounded, and the actual parts of a whole are limited. If there were no infinity (*apeirias*), all magnitudes would be commensurable

and there would be nothing inexpressible (*arrêton*) or irrational (*alogon*) features that are thought to distinguish geometry from arithmetic. (6. 7–22; transl. by Morrow 1970)

Proclus' terminology is almost the same as Plato's in the passages quoted from the *Philebus*. *Peras* and *apeiron* are the principles of mathematical entities and not of all things.<sup>22</sup> Ratios are "controlled by the principle of *peras*" just as in Plato. Like Plato, he relates *apeiron* and incommensurability and thinks that incommensurability is what distinguishes geometry from arithmetic. Immediately after (7. 1–5), Proclus says that without *peras* "there would be no commensurability or identity of ratios in mathematics, no similarity and equality of figures"; also that without *peras* sciences or arts are impossible. Proclus, speaking of magnitudes, refers also to infinite divisibility. It is not certain from this passage which is the relation between infinite divisibility and incommensurability. But in another passage (60. 11–16) he says that "in geometry a least magnitude has no place at all. Peculiar to geometry are the propositions regarding position. . . , the propositions about contacts—for contacts occur only when there are continuous magnitudes—and the propositions about the irrationals—for the irrational has a place only where infinite divisibility is possible." Hence, according to Proclus, in geometry we do not have a least magnitude or atoms but we have the continuous magnitudes. He thinks that the continuous and the discrete are mutually exclusive notions. But when Proclus later says that the irrational occurs only where infinite divisibility is possible and also that irrationality appears only in magnitude, he seems, like Aristotle, to identify magnitudes, continuity, and infinite divisibility. He does not see the possibility of an infinite divisibility that does not involve incommensurability, something that Plato probably saw.<sup>23</sup>

It is very probable that the Zenonian arguments and infinite divisibility did not play any significant role in ancient mathematics. The difficulties of Zeno's paradoxes do not have to do with mathematics in particular, but with philosophical and physical implications.<sup>24</sup> Geometers take for granted that they can cut a line at any place they want and can take a line as long as they want. Their problem was incommensurability. Incommensurability is the characteristic that distinguishes numbers from magnitudes and arithmetic from geometry.

Plato's approach to the continuum and magnitudes in the *Philebus* is more mathematical than physical. It seems that Plato considers incommensurability as an essential feature of magnitudes. According to my interpretation, we find in the *Philebus* an effort to explore the relation between continuous, infinite divisibility and incommensurability in contrast with commensurate things that are capable of appearing in ratios and proportions. It seems to me that Plato was conscious of the mathematical techniques dealing with incommensurable magnitudes and approaches the problem from this point of view.

## Notes

1. Of course such a distinction is possible only after the discovery of incommensurability, which probably happened in the last third of the fifth century BC, and made evident that magnitudes cannot be represented by numbers.



2. Translation is taken from Ackrill (1963). According to Aristotle, language (*phônê*) is discrete quantity because “it is measured by long and short syllables” (4b33–34).
3. See *Physics*, Book VI, chapters 1 and 2. At 232b24–5, Aristotle defines the continuum as “that which is divisible into divisibles that are always divisible” (cf. 185b10). Similarly, at 200b17–20, he says: “Infinite is presented firstly in the continuum; because of that, those who define the continuum happen to use the account of the infinite.” Modern mathematicians define continuity, more or less, in the same way (see Beth 1959, pp. 140–1; Anapolitanos 1985, p. 69). For a good examination of Aristotle’s account of continuity see White 1992, Part I.
4. See, e.g., *Physics*, Book Z, ch. 9. I shall call “Zenonian infinite divisibility” of a magnitude any kind of infinite divisibility according to an arithmetical ratio a:b. In this sense, the Aristotelian continuum is equivalent to what we nowadays call the set of rational numbers. But rational numbers are not equivalent to magnitudes, which are represented by real numbers. The set of rational numbers is continuous but not dense. The set of real numbers is both continuous and dense. The two sets are not equivalent. The first is a denumerable infinite set while the second is a super-denumerable infinite set. If in the set of rational numbers we add the set of irrational numbers, then we obtain the set of real numbers (see Anapolitanos 1985, p. 70; Sorabji 1983, p. 341).
5. Especially the incommensurability of the side and the diagonal of a square, which is one of Aristotle’s favorite examples: see e.g., *Met* 983a16, 1019b26, 1051b20; *Topics* 106b1, 163a12; *Prior Analytics* 41a28–9, 46b29–34; *On the Soul* 430a31; *NE* 1112a23. In the *Physics* we have only two references to the side-diagonal incommensurability (in book IV), but not in relation to the continuum or to infinite divisibility.
6. Of course, we cannot blame Aristotle for that. Only after Cantor’s treatment of infinity was it possible to deal adequately with these problems. There is, however, a passage (*On Generation and Corruption* 316a 15–317a13) in which Aristotle seems to doubt whether Zenonian infinite divisibility is enough to describe magnitudes. In this passage he tries to answer the difficulty (*aporia*) that if we accept that a body (i.e., a magnitude) is divisible throughout or everywhere (*pantêi diaireton*), “what will the body be which escapes division?” (316a15–7). Aristotle argues that division everywhere simultaneously of magnitudes is impossible but they are divisible anywhere (*opê(i)oun*). Of course, in this sense a magnitude can be divided not only in arithmetical ratios but also in others that cut the magnitude into incommensurable segments. Nevertheless, Aristotle does not draw this conclusion or speak about incommensurability in this passage.
7. Since my purpose is to understand the meaning and the characteristics of the terms *peras* and *apeiron* I am not going to translate them because any translation is also an interpretation.
8. The method consists in positing always

a single form (*idea*) in respect to everyone and search for it. . . and if we are successful, then after the one we should look for two, if there are two, or otherwise for three or whatever the number is; each of these ones should be treated in the same way, until one can see of the original one not only that it is one, a plurality, and an indefinite number, but also its precise quantity. But one should not attribute the character of indeterminate to the plurality until one can see the complete number between the indeterminate and the one. Then one can consign every one of them to the indeterminate with a clear conscience. (16d1–e2)

This method seems to be the method of division and collection and is presented as a method suitable for arts. The gift is given to people from the gods by Prometheus, and the examples that Plato gives later in order to clarify the method refer to the arts of music and grammar.

Translations from the *Philebus* are taken or adapted from Gosling (1975).

9. Regarding the *apeiron* in the passage 16ff we have two main different interpretations (see Frede 1992, pp. 427–8). According to the first interpretation, *apeiron* refers to the unlimited particular instances of a Form or a subject-matter (see Hackforth 1945, pp. 20–1; Benitez 1989,



- p. 53; Ross 1951, pp. 131–2). In this way, *apeiron* consists of an unlimited number of discrete particular things. According to the second interpretation *apeiron* refers to unlimitedly many “insignificantly different subkinds” of a kind (Crombie 1963, vol. II, p. 364) or to continua of opposite qualities, like the hotter and colder (Gosling 1975, pp. 169–70; Shiner 1974, p. 40; Moravcsik 1979, pp. 81–2; Barker 1996, p. 157; Meinwald 1998, pp. 169–70). Striker 1970, pp. 58, 80–1) is of the opinion that the term *apeiron* has quite different meaning in the two passages, although Plato (23c9–11) suggests the opposite. Some scholars interpret these passages in the *Philebus* in the light of the Aristotelian evidence about the Monad and the Indefinite Dyad (see e.g., Sayre 1983, pp. 149–55; Maziarz & Greenwood 1968, pp. 117–20).
10. I do not agree with Striker (1970, p. 45), who believes that “hotter and colder” are two *apeira* and not one. In that case we should have a demarcation point between hot and cold.
  11. The expression “more and less” (*mallon kai êtton*) appears only twice in other Platonic dialogues (*Protagoras* 356a4, *Timaeus* 87a3). Although the expression “more or less” (*mallon ê êtton*) is quite common, it is probable that the former expression was coined by Plato with technical meaning. The same expression appears in Aristotle.
  12. “Everything we find that can become more and less, and admits of strength and mildness, too much, and everything of that sort, we are to put in the category of the indeterminate (*apeiron genos*), as constituting a single class.”
  13. According to these characteristics Plato’s *apeiron* seems to be something totally indeterminate.
  14. Of course numbers or other discrete things of the same kind are always commensurable.
  15. Euclid’s definition of commensurable and incommensurable magnitudes is the following (Def. X, 1): “Those magnitudes are said to be commensurable which are measured by the same measure, and those incommensurable which cannot have any common measure.” But Proclus in his comments on the First Book of Euclid’s *Elements* says that “this is why both sciences (i.e., arithmetic and geometry) define commensurable magnitudes as those which have to one another the ratio of a number to a number, and this implies that commensurability exists primarily in numbers” (60.28–61.3; transl. by Morrow 1970).
  16. That Theaetetus worked on the theory of incommensurability is well testified by Plato himself (*Theaetetus* 147d–148e). The commentary on Euclid X (survived in Arabic and attributed to Pappus who takes information from Eudemus) says that most of the theory of irrationals found in Book X is work of Theaetetus (see Heath 1921, I 209). Moreover, an ancient scholiast on Euclid X, 9 definitely attributes the discovery of this theorem to Theaetetus (see Heiberg and Stamatis, 1977, Part 2, 113).
  17. In these two passages Plato speaks of how the mixed is produced by the introduction of *peras* into *apeiron*. When Plato speaks of *apeiron* in the framework of the mixed, it seems that he broadens or changes the meaning of it. Now *peras* gives not just any determination in *apeiron* but the right one according to the specific case or art. For example, if health is the right formula or balance between hot, cold, wet and dry, then other situations or formulas where we have excess of some elements and not equilibrium are considered as *apeiron*. I am not going to discuss this problem now because my aim is to show Plato’s influence from mathematics and especially from the phenomenon of incommensurability in his conception of *apeiron*.
  18. People who believe that Plato’s *apeiron* refers to continua (see n. 11) have not observed this serious problem.
  19. Many of the examples that Plato gives describing the mixed refer again to arts. He also devotes a very long passage (55d–57e) to the various arts and sciences. For the relevance of arts in the *Philebus*, see Gosling 1975, pp. 153–4, 169–72.
  20. Aristotle—who calls it *antanairesis*—refers to it at *Topics* 158b33–5.
  21. Probably it is difficult to understand how an infinite *anthuphairesis* process could be in practice a criterion for finding incommensurable magnitudes. We do not have clear examples where *anthuphairesis* was used in order to prove the incommensurability between two magnitudes. However this may happen in special cases where a geometrical figure is repeated infinitely in different sizes and segments (sides, diagonals etc.) in each new figure are produced by subtracting two other segments or when we have infinite repetition of the same gnomon. For

example, the diagonals of every regular pentagon produce another (smaller) regular pentagon, and so on, ad infinitum. Now if the side and the diagonal of the first pentagon are  $S_1$  and  $D_1$  ( $D_1 > S_1$ ), the diagonal of the second pentagon ( $D_2$ ) is  $D_1 - S_1$ ; and the side of the second pentagon ( $S_2$ ) is  $S_1 - D_2$ . So all sides ( $S_i$ ) and diagonals ( $D_i$ ) of these infinite pentagons are produced via an anthuphoretic process starting with the initial  $S_1$  and  $D_1$ . In this way we can conclude that the side and the diagonal of every regular pentagon are incommensurate. For other similar examples and more generally the technique of *anthuphairesis* in Greek mathematics, see Fowler (1987, chs. 2 and 5), Knorr (1975, chs. 2, 4, 7), and Sinnige (1968, pp. 73–80).

22. Of course, Proclus speaks here exclusively for mathematics.
23. For a good analysis of the notions of incommensurability and infinite divisibility in Proclus, see Anapolitanos and Demis (1996).
24. See Knorr (1975, pp. 116–125) who argues convincingly for that.

## Bibliography

- Ackrill, J. 1963. *Aristotle's Categories and De Interpretatione: Translated with Notes and Glossary*. Oxford: Oxford University Press.
- Anapolitanos, D. 1985. *Philosophy of Mathematics* (in Greek). Athens: Nepheli.
- Anapolitanos, D., and Á. Demis. 1996. "Incommensurability and infinity according to Proclus" (in Greek). *NEUSIS*, 5.
- Barker, A. 1996. "Plato's *Philebus*: The Numbering of a Unity." In *Dialogues with Plato*, edited by E. Benitez. Edmonton: Academic Printing & Publishing.
- Benitez, E.E. 1989. *Forms on Plato's Philebus*. Assen: Van Gorgum.
- Beth, E.W. 1959. *The Foundations of Mathematics*. Amsterdam: North-Holland Publishing Company.
- Crombie, I.M. 1963. *An Examination of Plato's Doctrines*. London: Routledge & Kegan Paul.
- Düring, I. 1991. *Aristoteles: Darstellung und Interpretation seines Denkens* (Greek transl.), Vol. A. Athens: Cultural Foundation of the Greek National Bank.
- Düring, I. 1996. *Darstellung und Interpretation seines Denkens* (Greek transl.), Vol. B. Athens: Cultural Foundation of the Greek National Bank.
- Fowler, D. 1987. *The Mathematics of Plato's Academy*. Oxford: Oxford University Press.
- Frede, D. 1992. "Disintegration and Restoration: Pleasure and Pain in Plato's *Philebus*." In *The Cambridge Companion to Plato*, edited by R. Kraut. Cambridge: Cambridge University Press.
- Gosling, J. 1975. *Plato: Philebus*. Oxford: Oxford University Press.
- Hackforth, R. 1945. *Plato's Philebus*. Cambridge: Cambridge University Press.
- Heath, T. 1921. *A History of Greek Mathematics*. Oxford: Oxford University Press.
- Heath, T. 1956. *Euclid: The Thirteen Books of the Elements* (3 vols). New York: Dover.
- Heiberg, I.L., and E.S. Stamatis, eds., 1977. *Euclidis Elementa* (Vol. V). Leipzig: Teubner.
- Karasmanis, V. 1990. "The Hypotheses of Mathematics in Plato's *Republic* and His Contribution to the Axiomatization of Geometry." In *Greek Studies in the Philosophy and History of Science*, edited by P. Nicolacopoulos. Dordrecht: Kluwer.
- Knorr, W.R. 1975. *The Evolution of the Euclidean Elements*. Dordrecht: Reidel.
- Knorr, W.R. 1982. "Infinity and Continuity: The Interaction of Mathematics and Philosophy in Antiquity." In *Infinity and Continuity in Ancient and Medieval Thought*, edited by N. Kretzmann. Ithaca: Cornell University Press.
- Maziarz, E., and T. Greenwood. 1968. *Greek Mathematical Philosophy*. New York: Frederick Ungar.
- Meinwald, C.C. 1998. "Prometheus's Bounds: *Peras* and *Apeiron* in Plato's *Philebus*." In *Method in Ancient Philosophy*, edited by J. Gentzler. Oxford: Oxford University Press.
- Moravcsik, J.M. 1979. "Forms, Nature, and the Good in the '*Philebus*'." *Phronesis* 24(1):81–104.

- Morrow, G.R., ed., trans. 1970. *Proclus: A Commentary on the First Book of Euclid's Elements*. Princeton: Princeton University Press.
- Proclus 1970. *A Commentary on the First Book of Euclid's Elements*. Edited and translated by G.R. Morrow. Princeton: Princeton University Press.
- Ross, D. 1951. *Plato's Theory of Ideas*. Oxford: Oxford University Press.
- Sayre, K. 1983. *Plato's Late Ontology*. Princeton: Princeton University Press.
- Shiner, R. 1974. *Knowledge and Reality in Plato's Philebus*. Assen: Van Gorgum.
- Sinnige, T.G. 1968. *Matter and Infinity in the Presocratic Schools and Plato*. Assen: Van Gorgum.
- Sorabji, R. 1983. *Time, Creation and the Continuum*. London: Duckworth.
- Striker, G. 1970. *Peras und Apeiron*. Göttingen: Vandenhoeck & Ruprecht.
- White, M. 1992. *The Continuum and the Discrete*. Oxford: Oxford University Press.

# The Works of Gerasimos Santas

## Books

- 1979. *Socrates*. London: Routledge.
- 1988. *Plato and Freud: Two Theories of Love*. Oxford: Blackwell.
- 1990. Italian translation of 1988, *Plato e Freud*. Bologna: il Mulino.
- 1997. Greek translation of 1979, *Σωκράτης*. Athens: Ellinika Grammata.
- 2001. *Goodness and Justice: Plato, Aristotle and the Moderns*. Oxford: Blackwell.
- 2003. Italian translation of 1979, *Socrate*. Milano: V&P Universita.
- 2008. *The Blackwell Guide to Plato's Republic* (ed.). Oxford: Blackwell.
- 2010. *Understanding Plato's Republic*. Oxford: Wiley-Blackwell.
- 2011. Greek translation of 2001, *Το Αγαθόν και η Δικαιοσύνη*. Athens: Kardamitsas.

## Articles

- 1964. "The Socratic Paradoxes." *The Philosophical Review* 73.
- 1965. Reprinted in A. Sesonske and E. Fleming (eds.), *Plato's Meno*. Belmont: Wadsworth Publishing Company.
- 1966. "Plato's *Protagoras* and Explanations of Weakness." *The Philosophical Review*, 75.
- 1967. Reprinted in *Reprint Series in Philosophy*. New York: Bobbs- Merrill.
- 1969. Reprinted in G. Mortimer (ed.), *Weakness of Will*. London: Macmillan.
- 1969. "Socrates at Work on Virtue and Knowledge in Plato's *Laches*." *Review of Metaphysics* 22.
- 1969. "Aristotle on the Explanation of Action, Practical Inference, and *Akrasia*." *Phronesis* 14.
- 1971. Reprinted in G. Vlastos (ed.), *The Philosophy of Socrates*. Garden City: Doubleday Anchor Books.
- 1972. "The Socratic Fallacy." *The Journal of the History of Philosophy* 10.
- 1973. "Hintikka on Knowledge and Its Objects in Plato." In *Patterns in Plato's Thought*, edited by J.M.E. Moravcsik. Dordrecht: Reidel.
- 1973. "Socrates at Work on Virtue and Knowledge in Plato's *Charmides*." In *Exegesis and Argument*, edited by E.N. Lee et al. Assen: Supplementary Volume of *Phronesis*.
- 1979. "Plato's Theory of Eros in the *Symposium*." *Nous* 13.
- 1980. "The Form of the Good in Plato's *Republic*." *Philosophical Inquiry* 2.
- 1981. "Plato on Love, Beauty and the Good." In *The Greeks and the Good Life*, edited by D. Depew. Fullerton: A Publication of California State University, Fullerton.
- 1982. "Passionate Platonic Love in the *Phaedrus*." *Ancient Philosophy* 2.
- 1983. Reprinted in J. Anton and A. Preus (eds.), *Essays in Ancient Philosophy*, Vol. II. Albany: State University of New York Press.

- 1984. "Plato on Friendship and Familial Love in the *Lysis* and the *Republic*." *Philosophical Inquiry* 6.
- 1985. "Two Theories of Good in Plato's *Republic*." *Archiv für Geschichte der Philosophie* 57.
- 1986. "Plato on Goodness and Rationality." *Revue Internationale de Philosophie* 40.
- 1988. "Justice and Democracy in Plato's *Republic*." In *Antike Rechts und Sozialphilosophie*, edited by O. Gigon and M. Fischer. Bern: Peter Lang.
- 1989. "Desire and Perfection in Aristotle's Theory of the Good." *Apeiron* 22.
- 1989. "Aristotle's Criticism of Plato's Theory of the Good: Ethics Without Metaphysics?" *Philosophical Papers* 18.
- 1990. "Knowledge and Belief in Plato's *Republic*." In *Greek Studies in the Philosophy and History of Science*, edited by P. Nicolacopoulos. Boston: Kluwer Academic Publishers.
- 1992. "The Theory of Eros in Socrates's Second Speech." In *Understanding the Phaedrus*, edited by L. Rossetti. Sankt Augustin: Academia Verlag.
- 1992. "Socratic Goods and Socratic Happiness." In *The Philosophy of Socrates*, edited by K. Voudouris. Athens: International Center for Greek Philosophy and Culture.
- 1992. "Did Aristotle Have a Virtue Ethics?" *Philosophical Inquiry* 14.
- Reprinted in D. Statman (ed.), *Virtue Ethics*. Washington: Georgetown University Press.
- 1993. Reprinted in T. Irwin and M. Nussbaum (eds.), *Virtue, Love, and Form*. Edmonton: Academic Printing and Publishing.
- 1995. "The Relation Between Aristotle's Ethics and Politics." In *Aristotelian Political Philosophy*, edited by K. Voudouris. Athens: International Center for Greek Philosophy and Culture.
- 1996. "The Structure of Aristotle's Ethics." *Topoi* 15.
- 1996. "Goodness: Aristotle and the Moderns, A Sketch." *Philosophical Inquiry* 18.
- 1997. Reprinted in W.J. Prior (ed.), *Socrates: Critical Assessments*. London: Routledge.
- 1998. "Plato and Freud on Justice and Health." In *Philosophy and Medicine*, edited by K. Voudouris. Athens: International Association for Greek Philosophy.
- 1999. Reprinted in G. Fine (ed.), *Plato 1*. Oxford: Oxford University Press.
- 2000. "Plato's Criticism of the Democratic Man in the *Republic*." *The Journal of Ethics* 5.
- 2002. Translated into Greek and reprinted in *Πλάτων Vol. II: Essays in Honor of D. Z. Andriopoulos*. Athens: Papademas.
- 2002. "Plato: Ethics." In *The Blackwell Guide to Ancient Philosophy*, edited by C. Shields. Oxford: Blackwell.
- 2003. "The Form of the Good as Paradigm and Its Essence." In *New Images of Plato*, edited by G. Reale and S. Scolnicov. Sankt Augustin: Academia Verlag.
- 2003. "Penner, Plato, and Sachs on Justice and Happiness." In *Desire, Identity, and Existence*, edited by N. Reshotko. Kelowna: Academic Printing and Publishing.
- 2003. "Gender and Justice in Plato's *Republic* and *Laws*." In *Proceedings of the VI Symposium Platonicum*. Sankt Augustin: Academia Verlag.
- 2004. "Legalite, Justice et Femmes dans la *Republic* at les *Lois* de Platon." *Revue Francaise d'Histoire des Idees Politique* 16.
- 2005. "Plato." In *Meet the Philosophers of Ancient Greece*, edited by P. O. Grady. Burlington: Ashgate.
- 2006. "Justice, Law, and Women in Plato's *Republic*." In *Law and Rights in the Ancient Greek Tradition*, edited by G. Anagnostopoulos. Athens: Supplementary Volume of *Philosophical Inquiry*.
- 2006. "Methods of Reasoning about Justice in Plato's *Republic*." In *The Blackwell Guide to Plato's Republic*, edited by G. Santas. Oxford: Blackwell.
- 2006. "Plato on Pleasure as the Human Good." In *The Blackwell Companion to Plato*, edited by H. Benson. Oxford: Blackwell.
- 2007. "Plato's Criticisms of Democracy in the *Republic*." In *Freedom, Reason and Polis: Essays in Ancient Greek Political Philosophy*, edited by D. Keyt and F.D. Miller, Jr. Cambridge: Cambridge University Press.

- 2009. “The Socratic Method in *Republic I*.” *Skepsis XX*.
- 2010. “The Socratic Method and Ethics.” In *Foundations of Ancient Ethics*, edited by J. H. Hardy and G. Rudebusch. Göttingen: Vandenhoeck.
- 2011 “The Sophists.” In *International Encyclopedia of Ethics*. Oxford: Wiley-Blackwell.
- 2011. “The Isomorphism Between the Just City and the Just Person in Plato’s *Republic*.” forthcoming.

# Index

## A

- Abilities, 206, 208–215, 232–233, 236, 251, 343, 385  
    human, 215  
    intellectual, 213
- Ackrill, J.L., 309, 335–337, 340, 345, 347, 352, 359, 362
- Action, 3, 8–12, 25, 47, 51, 78, 99, 105, 108–110, 112–120, 136–141, 147–149, 151, 153–155, 197, 199, 200, 202, 227, 252, 263, 265–268, 331–332, 338, 340, 342, 344, 349, 356–357, 360, 363, 368–371, 384, 387  
    actual, 113–114, 137–141  
    non-virtuous, 6, 9  
    particular, 8, 25, 108, 114, 117–118, 136, 139, 332  
    theory of, 147–149, 369
- Adam, J., 188, 274
- Adeimantus, 186–188, 194, 208, 211–213, 229–232, 255
- Agathon*, 2, 95, 154–155, 167, 170–171, 190, 201, 276, 292, 298–299, 301–302, 304–310, 312–314, 335–337, 341–342, 344–345, 356, 377  
    *See also* Good/goodness
- Aitia*, 232, 235–236, 286–287, 298–302, 305–306, 308–309, 311–312, 314, 316, 390
- Anagnostopoulos, M., 367–385
- Anaxagoras, 2, 14, 301, 304, 310, 314
- Annas, J., 259–260, 262, 312
- Anytus, 224–225, 231, 236
- Apeiron*, 390–395
- Appetite, 159, 252–253, 264, 266, 268–269, 324–325, 329, 372, 385
- Aristocracy, 195–201, 262
- Aristotle

Function Argument, xxvii, 335–358  
works

- Categories*, 389–390  
*de Anima*, 330, 370–371, 381  
*Eudemian Ethics*, 370, 378  
*On Generation and Corruption*, 396  
*Metaphysics*, 11, 146, 372, 376  
*Movement of Animals*, 330  
*Nicomachean Ethics*, 202, 361, 367–368, 377, 383–384  
*Physics*, 356, 389–390, 392  
*Politics*, 202, 377  
*Posterior Analytics*, 57  
*Prior Analytics*, 389  
*Rhetoric*, 72, 376  
*Topics*, 57, 61–62, 287–288, 368, 371–373

## B

- Benson, H.H., 21, 23–24, 219–237
- Bolton, R., 311–312
- Boulêsis*, 96, 98–99, 324, 329, 368, 372, 380
- Bravery, 80, 213, 258, 261–262, 265, 268  
    *See also* Courage
- Brickhouse, T.C., 93–101
- Brisson, L., 145
- Burnet, J., 359
- Burnyeat, M.F., 273

## C

- Callicles, 31, 42, 44, 60, 77–78, 81, 86, 149, 152–156, 172, 323–324, 328
- Carpenter, M., 48
- Cause, 12, 67, 78, 83, 187, 197, 200, 206, 222, 232–233, 235–237, 269, 276, 287, 289, 298–299, 301, 303–306, 308–316, 327, 331, 348, 355–357, 381, 389–390  
    causal explanation, 297–316



- Cause (*cont.*)  
 causal relations, 300, 302, 308,  
 310–311, 313  
 definition of, 315–316  
 formal cause, 287, 289, 308, 314,  
 356–357
- Cleinias, 62, 71
- Cooper, J.M., 149, 364
- Courage, 2, 5, 7–8, 25–26, 29, 62, 80, 147,  
 165–166, 168, 175–176, 189,  
 228, 250  
*See also* Bravery
- D**
- Dancy, R.M., 70–72
- Democracy, 187, 190, 192–202, 247, 260–261,  
 268, 327  
 actual, 187, 195  
 democratic rule, 192–193, 200
- Desire  
 for the apparent good, 106, 113, 147, 332,  
 367–376  
 de dicto, 107, 125–127, 129–135,  
 141–142  
 de re, 107, 125–142  
 dominant, 326–327  
 for the real good, 106, 108, 120, 147,  
 367–385
- Donnellan, K., 137
- E**
- Edelstein, L., 89–90
- Education, 59, 77, 80–81, 85, 147, 189–190,  
 193, 200–201, 209, 214, 228,  
 232–233, 236, 249, 251, 253, 268,  
 271, 282, 284–285, 302–303
- Elenchus* (*elenchos*), 11–12, 19–31, 35–46,  
 54–56, 58–60, 65, 69, 96, 194, 221  
 elenctic method, xxiv, 12, 35–36, 40–43,  
 45–46, 67
- Emotions, 167–168, 329–330
- Epagôgê*, 54–58, 60–63, 65–69, 151,  
 153–154  
 explicative, 54, 60, 62, 69  
 theory of, 63
- Epistemology, 22, 146, 282, 307–308
- Epithumein*, 95, 97–98
- Eros, 83, 191–192, 198–199, 202, 327
- Euclid, 308, 392–394
- Eudaimonia*, 332, 335, 340, 345–346, 349,  
 352, 354, 370  
 nature of, 116, 370, 378
- Excellences, 54, 147, 186–187, 190, 192,  
 196, 199, 207, 209, 250, 258,  
 272, 335–336, 338–344, 356, 374,  
 378–379, 383
- Experts, 31, 64–65, 126, 171
- F**
- Food, 97, 110, 127, 178, 253, 259, 325–326,  
 328, 330–332, 381
- Frege, G., 103–122, 140–141
- Freud, S., xxv
- G**
- Gill, C., 90, 92
- Glassen, P., 335–338, 340, 342–344, 349,  
 357, 361
- Glaucon, 187–188, 214–215, 225–227, 234,  
 245, 247–248, 250, 255, 272–274,  
 281–283
- Good/goodness, 55, 60, 145, 167, 171,  
 175–176, 178, 190, 205–206,  
 245, 251, 258–259, 271, 281–293,  
 313–314, 335–337, 341–345, 349,  
 352, 361, 372, 374–375, 377–378,  
 381–385  
*See also* *Agathon*; Human good
- Guthrie, W.K.C., 55, 58, 188, 190, 197
- H**
- Hierarchy, 110, 112, 115, 129, 134–136, 139,  
 260, 289  
 hierarchical structure, 127, 130–131  
 means-end, 110, 112, 115  
 natural, 260
- Hobbes, 215, 316
- Human good, 107, 134, 165, 171, 206, 248,  
 323, 335–342, 344–346, 348–349,  
 353–355, 357, 378, 383
- I**
- Ignorance, 9, 21–22, 59, 81, 93, 99, 108,  
 111, 147, 167, 170–178, 219, 221,  
 237, 312
- Injustice, *see* Justice  
 suffering of, 153, 155
- Instrumental value, 83, 151, 253, 346, 348,  
 350–356, 362, 370
- Irwin, T., 94, 149, 151–155, 251, 369,  
 373–374, 379–380

**J**

- Joseph, H.W.B., 283, 291–293  
 Justice, 2, 5–7, 60–61, 80, 82, 93, 147, 151, 154, 166–167, 169, 171, 176–178, 185–202, 205, 207, 213, 215, 223, 228–229, 234–235, 245–254, 255–270, 271–276, 281, 288–289, 299, 306, 315, 358  
 account of, 167, 246, 248, 271  
 definition of, 25, 185–202, 254, 271  
 and human happiness, 271–276  
 injustice, 61, 77, 93, 149–153, 155–156, 186–187, 190, 192, 195, 199–202, 234, 247–248, 250, 255, 261, 267, 325  
 theory of, 186, 255–256, 262

**K**

- Kallipolis, 208, 220, 223, 225–226, 228, 236, 258–262, 268–269  
 Kenny, A., 337, 340, 345, 347–350, 352  
 Keyt, D., 255–270, 288, 307  
 Knowledge, 2–4, 6–13, 21, 55–59, 61, 63–65, 68–69, 76, 79, 84, 88, 93, 100, 107–108, 119, 127, 131, 147–148, 156, 165–169, 171–175, 179, 189–191, 197–198, 211–212, 219–237, 261, 281–287, 291, 298–299, 301, 303–306, 308–309, 312–314, 364, 370, 373, 376, 379, 384–385  
 comprehensive, 107, 122  
 ethical, 8–11  
 virtue is, 19, 93, 119, 147, 165, 219–220, 224, 226, 228, 230, 235–236  
 Kraut, R., 36, 38–40, 46, 58, 271  
 Krohn, A., 282–283, 291

**L**

- Locke, J., 215, 217, 316  
*Logistikon*, 187, 189–190, 199–201, 325  
*Logoi*, 54, 75, 77–81, 304, 394  
*Logos*, 76–79, 81–84, 86–87, 105, 187, 190, 194, 197–198, 200, 202, 223, 230, 247, 253, 269  
 Love, xxv–xxvi, 68, 83–84, 121, 167, 191, 208–212, 214–215, 228, 251, 253, 264, 275–276, 307, 327–328, 377

**M**

- Marriage, 125, 128, 130–131, 234  
 Mathematics, 314, 389–395  
 McLaughlin, P., 361–362

- McPherran, M., 53–69  
 Medicine, 58, 60, 171, 356  
 Morgan, K.A., 89–90, 92  
 Myth, 76–88, 257, 260–261  
 educational, 77, 80–81, 85  
*muthologikos*, 75–76, 85, 88  
*muthos*, 76–79, 81, 83, 85–88, 103–105  
 myth of Er, 80–82, 86, 188  
 philosophical, 77, 81, 86, 88, 189

**N**

- Nettleship, R.L., 203–204, 237, 240–242, 293, 296  
 Nicias, 8, 166, 168–169, 172, 174–176, 178–179

**O**

- Oligarchy, 195–196, 199, 268, 326–327  
*Ousia*, 83, 87, 283–284, 286, 290–291, 311

**P**

- Pain, 97, 155, 159, 200, 269, 324, 328–329, 337  
 Pedagogy, 165–179  
 Penner, T., xxiv, 94, 103–122, 126, 132–133, 135–137, 146, 367–368, 370, 376  
 Piety, 2–3, 8, 10–12, 25, 59, 67, 165–166, 168–169, 171, 177–179, 196

**Plato**

- Forms, 206, 390  
 Function Argument, 258, 387  
 Plato's Academy, 393  
 works  
*Alcibiades*, 241, 358–359  
*Apology*, 59, 75, 165, 167–171, 175, 187, 191–192, 194, 202  
*Charmides*, 5, 7–10, 56, 62–63, 65, 68, 98, 165, 167  
*Crito*, 36, 42, 53, 59–60, 64–65, 147, 187, 269–270  
*Euthyphro*, 2–4, 6, 8–12, 25, 59–60, 67, 165–166, 169, 174–178, 187, 350  
*Gorgias*, 21, 30–31, 36, 41–42, 44, 46, 57, 59–60, 62, 76–79, 81, 85–86, 93, 106, 108–109, 112–114, 116, 118, 120, 145–157, 172, 188, 225, 323–325, 328, 363, 367–368, 384  
*Hippias Major*, 31, 237, 318  
*Hippias Minor*, 71, 237  
*Laches*, 5, 7–8, 13, 25, 31, 42, 63, 165–167, 169, 174–179  
*Laws*, 256–257, 268–269, 312  
*Lysis*, 105, 146, 150, 165, 167, 170

Plato (*cont.*)

- Meno*, 13, 25, 30, 55, 57, 61, 67, 93–101, 103–106, 108–112, 114, 118–120, 166–170, 173–179, 219–230, 233, 235–237, 305, 310
- Parmenides*, 255, 257
- Phaedo*, 2, 75–81, 83–87, 146, 185, 187–188, 195, 206, 220–224, 234–236, 297, 312
- Phaedrus*, xxv, 76–77, 83–88, 167, 189, 191, 253, 257, 312
- Philebus*, 390, 393–395
- Politicus (Statesman)*, 256, 262, 269, 360
- Protagoras*, 7, 13, 31, 60, 62, 100, 146, 165, 167, 169–172, 188, 257, 372, 375–378
- Republic (selected)*, xxv–xxviii, 196, 205–215, 245–254, 281–293, 307
- Symposium*, xxv, 191
- Theaetetus*, 119, 256–257, 393
- Timaeus*, 77, 86–89, 167, 189, 196, 255, 286, 289, 312
- Pleonexia*, 210, 262–263
- Poetry, 75–76, 78, 82, 86, 88, 269
- Polansky, R., 48
- Polemarchus, 15, 49, 188, 194, 216, 245–246, 250, 266
- Polis*, 186–203, 231, 256–265, 267–269
  - good of, 197, 262
  - ideal, 259, 269
- Polus, 42–43, 148–156, 172, 323, 368
- Prison, 47–48, 75–76, 108, 185, 300–301, 306, 314
- Properties, 79, 131, 139–140, 154, 157–158, 216, 228, 285, 288, 290, 303, 310–311, 313, 317, 319, 348, 364, 389
  - sensible, 157–158, 258
- Psyche, 64–65, 69, 258, 262–265, 267–269, 302, 312
  - psychic activity, 330
  - psychic faculties, 324–326, 329, 333
  - psychic justice, 258, 262–266, 268, 272–275
  - psychological states, 8, 104–105, 107, 125, 131

**R**

- Rackham, H., 373
- Rickless, S., 180

- Robinson, R., 54–57, 69–70, 304
- Rowe, C.J., 132–133, 137, 145–156, 367–370, 376

**S**

- Santas, G.X., 23, 26–27, 36, 39–42, 46, 53–69, 75, 93–97, 99–100, 111, 205–206, 215, 245–254, 258, 262, 269, 271–273, 288, 307–309, 313–316, 323–326, 329, 332–333, 383–385
- Science, 54, 61, 68, 171, 303, 363, 384
- Scott, D., 239
- Sensible objects, 307, 310, 313–315
- Sensible world, 197–198, 286, 292, 307, 311, 313, 390
- Simonides, 170, 172, 246
- Smith, N.D., 93–101
- Socrates
  - Historical, xxiii, 13, 31, 69–70, 89, 145, 276–277
  - Plato's Socrates, 76–77, 277, 384
- Sophists, 69–70, 224–225, 239
- Soul, 7, 76–79, 81–88, 147, 149, 165, 167, 169–170, 175, 186–187, 189–193, 197–202, 206–208, 210, 212–213, 222–223, 228–229, 250–254, 257–258, 264–269, 272–276, 285, 303, 305, 312, 324–333, 335–336, 342–344, 370, 372, 375, 378, 383–384
  - activity of, 335–336, 342–344, 384
  - divided, 149, 159
  - embodied, 197, 265, 269
  - individual, 189, 258
  - tripartite, 84, 167, 208
- Sufficiency thesis, 3–4, 8, 10, 14–16
- Szlezák, T., 91

**T**

- Technê*, 190, 231, 256, 259, 356, 393
- Thrasymachus, 15, 31, 33, 49, 60, 186, 206, 245–248, 250, 257, 267, 350
- Timocracy, 160, 196, 199, 203, 268
- Truth, 1, 11–12, 20, 22–25, 27–29, 35, 37–39, 41–44, 55–56, 58, 61, 64, 78, 81, 83–85, 87–88, 93, 109, 137, 172–173, 191–192, 211–213, 215, 220–224, 227–231, 236, 247, 249, 253, 256–257, 269, 286–287, 291, 293, 299, 305–306, 313, 323, 333, 337, 339, 344, 367, 376, 378–381, 383
  - criterion of, 23–24, 249

establishing, 41–43, 50  
 love of, 211–212, 228  
 Tyranny, 160, 196, 199–203, 213, 261,  
 268, 327

## V

Virtue, 1–13, 19, 25, 28, 54, 60, 65–67,  
 79–80, 84, 93–94, 96, 99, 119,  
 134, 147, 154, 165–169, 172,  
 174–179, 189–191, 193, 196–198,  
 200, 207–208, 211–213, 215,  
 219–237, 249, 258–259, 261–265,  
 268, 271–272, 274–275, 281–282,  
 285, 288–289, 309, 330–332,  
 340, 359–360, 362, 368, 374–378,  
 380–384  
 account of, 167, 181  
 actions of, 4, 6, 8–9, 14, 382–383  
 complete, 7, 16, 232, 240  
 human, 1, 49, 165–167, 181, 274–276  
 imperfect, 240  
 intellectualist conception of, 7–8,  
 11–13, 15  
 justice is, 166, 176  
 nature of, 1, 13, 219, 237, 242

possession of, 7, 332, 359, 383  
 teaching, 13, 225, 238  
 true cardinal, 212–213  
 Vlastos, G., 19–30, 35–40, 42–44, 53–54,  
 57–63, 66, 69, 94, 245, 262, 274,  
 288, 307, 311–314

## W

Whiting, J., 337, 345, 347–350, 352–353, 355,  
 359, 362–365  
 Wilkes, K., 335–337, 340, 345–347, 350, 352,  
 358–359, 362, 364  
 Wisdom, 2, 10, 57, 61–62, 79–80, 82, 88,  
 165–179, 185, 189, 191–192, 195,  
 198, 200, 208, 210–215, 229, 233,  
 236, 245, 250–251, 253, 258–259,  
 261–262, 264, 275–276, 284, 298,  
 328, 371, 375, 385  
 levels of attainment of, 166, 169, 171  
 lover of, 211–212, 251, 253, 328  
 real, 169  
 Woodruff, P., 180

## X

Xenophon, 54, 60, 69–70, 89, 103