

Synthese Library 372

Julie Zahle  
Finn Collin *Editors*

# Rethinking the Individualism- Holism Debate

Essays in the Philosophy of Social  
Science

 Springer

# Rethinking the Individualism-Holism Debate

# SYNTHESE LIBRARY

STUDIES IN EPISTEMOLOGY,  
LOGIC, METHODOLOGY, AND PHILOSOPHY OF SCIENCE

*Editor-in-Chief*

LUCIANO FLORIDI, University of Oxford, Oxford Internet Institute,  
United Kingdom

*Editors*

THEO A.F. KUIPERS, University of Groningen Fac. Philosophy, The Netherlands

TEDDY SEIDENFELD, Carnegie Mellon University Dept. Philosophy, USA

PATRICK SUPPES, Stanford University Ctr. Study of Language & Information,  
USA

JAN WOLEŃSKI, Jagiellonian University of Krakow Institute of Philosophy,  
Poland

DIRK VAN DALEN, Utrecht University Department of Philosophy,  
The Netherlands

VOLUME 372

For further volumes:  
<http://www.springer.com/series/6607>

Julie Zahle • Finn Collin  
Editors

# Rethinking the Individualism-Holism Debate

Essays in the Philosophy of Social Science

 Springer

*Editors*

Julie Zahle  
Department of Media, Cognition  
and Communication  
University of Copenhagen  
Copenhagen, Denmark

Finn Collin  
Department of Media, Cognition  
and Communication  
University of Copenhagen  
Copenhagen, Denmark

ISBN 978-3-319-05343-1

ISBN 978-3-319-05344-8 (eBook)

DOI 10.1007/978-3-319-05344-8

Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014937661

© Springer International Publishing Switzerland 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

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

# Contents

<b>1</b>	<b>Introduction</b> . . . . .	<b>1</b>
	Julie Zahle and Finn Collin	
<b>Part I Ontological Individualism-Holism</b>		
<b>2</b>	<b>What Is Individualism in Social Ontology? Ontological Individualism vs. Anchor Individualism</b> . . . . .	<b>17</b>
	Brian Epstein	
<b>3</b>	<b>Social Entities and the Basis of Their Powers</b> . . . . .	<b>39</b>
	Dave Elder-Vass	
<b>4</b>	<b>Actor-Centered Sociology and the New Pragmatism</b> . . . . .	<b>55</b>
	Daniel Little	
<b>5</b>	<b>Three Issues in Social Ontology</b> . . . . .	<b>77</b>
	Philip Pettit	
<b>6</b>	<b>Collective Responsibility and Group-Control</b> . . . . .	<b>97</b>
	András Szígeti	
<b>7</b>	<b>Rethinking Micro-Macro Relations</b> . . . . .	<b>117</b>
	Petri Ylikoski	
<b>Part II Methodological Individualism-Holism</b>		
<b>8</b>	<b>Dead Ends and Live Issues in the Individualism-Holism Debate</b> . . .	<b>139</b>
	Harold Kincaid	
<b>9</b>	<b>Explanatory Strategies Beyond the Individualism/Holism Debate</b> . . .	<b>153</b>
	Jeroen Van Bouwel	
<b>10</b>	<b>Holism, Emergence, and the Crucial Distinction</b> . . . . .	<b>177</b>
	Julie Zahle	

**11 Who Are the Agents? Actor Network Theory, Methodological Individualism, and Reduction . . . . . 197**  
Finn Collin

**12 Structure, Agency, and Improvisation . . . . . 219**  
Mark Risjord

**13 Are Individuals Fickle-Minded? . . . . . 237**  
Mathew D. McCubbins and Mark Turner

**Index . . . . . 253**

# Chapter 1

## Introduction

Julie Zahle and Finn Collin

**Abstract** The introduction provides an overview of the ontological and the methodological individualism-holism debates. Moreover, these debates are briefly discussed in relation to two kindred disputes: The micro-macro and the agency-structure debates. Finally, the contributions to this book are briefly presented.

The individualism-holism debate is an old – but still vibrant – dispute within the philosophy of the social sciences and the social sciences themselves. Over the course of its history, there are three phases in which the discussion has been particularly lively (Udehn 2002: 479). The first is around the turn of the nineteenth century with significant contributions by, among others, Emile Durkheim and Max Weber. The second phase is around the 1950s, where the ardent defenses of methodological individualism by Friedrich Hayek, Karl Popper, and J.W.N. Watkins spurred on the debate. Finally, the third and last phase stretches from the 1980s and up till today with pioneering contributions by a number of theorists including Roy Bhaskar, Raymond Boudon, James S. Coleman, Jon Elster, Alan Garfinkel, Daniel Little, Harold Kincaid, and Philip Pettit.<sup>1</sup>

The individualism-holism debate has first and foremost revolved around two issues:

- What is the ontological status of social phenomena and, as part of this, their relationship to individuals?

---

<sup>1</sup> It should be stressed that Udehn simply lists the three periods. Moreover, his claim that the third phase of the debate is still ongoing is from 2002. We are responsible for the observation that the third phase extends to the present just as the addition of key figures within all three periods is our doing.

J. Zahle (✉) • F. Collin  
Department of Media, Cognition and Communication, University of Copenhagen,  
Njalsgade 80, 2300, Copenhagen, Denmark  
e-mail: [jzahle@hum.ku.dk](mailto:jzahle@hum.ku.dk); [collin@hum.ku.dk](mailto:collin@hum.ku.dk)



- To what extent may, and should, social scientific explanations focus on individuals and social phenomena respectively?

The second question, in particular, has received a lot of attention. The *methodological* individualism-holism debate refers to the discussion of this issue, whereas the *ontological* individualism-holism issue denotes disputes relating to the first question. As part of these discussions, a number of other topics have been addressed too relating to meaning, confirmation, research heuristics, ethics, and the like. Accordingly, it is also possible to distinguish between individualism-holism debates on meaning, confirmation, research heuristics, and so on.

The present collection of papers is in line with the tradition in terms of its focus. The contributions divide into two categories: One group focuses primarily on the ontological dispute, the other on the methodological dispute. The papers reflect the newest trends within these debates in the following ways: First, many papers point to basic concepts and distinctions that are widely applied within the debate. Rather than being uncritically taken for granted, it is suggested, these concepts and distinctions are in need of being carefully spelled out, questioned, or even replaced by more adequate conceptions. This approach is, among other things, adopted in relation to the notion of supervenience, the notion of agency and its domain of application, the distinction between individuals and social phenomena, and the distinction between explanations that focus on individuals and social phenomena respectively. Second, many of the papers strike out in a new direction by paying close attention to actual developments within social scientific research. They explicitly state – and exemplify – a focus on issues raised by, or addressed within, the context of social scientific research, just as they make suggestions of relevance to social scientific practice. In these ways, then, the papers exemplify a rethinking of the debate that point to novel directions in which to take future philosophical discussions and future empirical work in the social sciences.

In the following, we provide a rough systematic overview of the individualism-holism debate. As we go along, we also comment on the history of debate. In Sect. 1.1, we outline the ontological dispute and, in Sect. 1.2, the methodological debate. Further, in Sect. 1.3, we discuss the individualism-holism debate in relation to two kindred disputes.<sup>2</sup> Against that background, we briefly present the contributions to this book.

## 1.1 The Ontological Individualism-Holism Debate

The ontological individualism-holism debate concerns the ontological status of social phenomena (or facts) and, as part of this, their relationship to individuals (or facts about individuals). Ontological holists contend that social phenomena

---

<sup>2</sup> Sections 1.1, 1.2, and 1.3 draw and expand on Zahle (2007, 2013).

exist *sui generis* or over and above individuals, whereas ontological individualists deny this. Before looking into various interpretations of the idea that social phenomena exist *sui generis*, it is instructive briefly to characterize the notions of social phenomena and individuals.

Within the individualism-holism debate more generally, it is common to distinguish between various kinds of social phenomena. Some of the most frequently mentioned ones are: (a) social organizations, as exemplified by a nation, a firm, and a university; (b) statistical properties like the literacy or suicide rate of a group of individuals; (c) norms and rules as illustrated by the rule to drive to the right and the prohibition against sex with close family members; (d) cultures such as the Mayan culture; and (e) social structures as typically identified with one or several of the sorts of social phenomena already listed. The ontological debate has mainly concentrated on the issue of whether social organizations exist over and above individuals. In this connection, social organizations are simply referred to by way of terms such as “university,” “hospital,” and the like.

It is less obvious how individuals should be characterized from the perspective of offering an ontological analysis of their relationship to social phenomena. Not surprisingly, therefore, opinions diverge on this question. Most notably, it is a matter of dispute exactly what kinds of interactions and interrelations between individuals it is permissible to invoke as part of ontological analyses of this sort. Also, and related to this, it is discussed whether it is admissible to mention individuals’ beliefs about social organizations like nations and firms. Or, to mention one last example, there are different views as to whether a specification of individuals should be taken broadly to include the objects they make use of and, more generally, the physical environment in which they find themselves.

Throughout the history of the ontological debate, a variety of suggestions have been made as to what it takes for social phenomena to exist *sui generis*. Here are some of the many, and sometimes overlapping, ways in which this idea has been fleshed out:

- The causal overriding criterion: Social phenomena such as social organizations exist over and above individuals insofar as they have causal powers that are independent of, and override, the causal powers of individuals.
- The translation criterion: Social phenomena such as social organizations exist over and above individuals insofar as terms, like “nation” and “school,” that refer to these phenomena cannot be translated into statements about individuals.
- The composition criterion: Social phenomena such as social organizations exist over and above individuals insofar as they are not *merely* composed of ensembles of individuals.
- The determination criterion: Social phenomena such as social organizations exist over and above individuals insofar as individuals do *not* non-causally determine what kinds of organizations, properties, and the like, are being instantiated.
- The agency criterion: Social phenomena such as social organizations exist over and above individuals insofar as they qualify as group agents that have attitudes supervenient upon the attitudes of individuals.

In his writings from around the turn of the nineteenth century, Durkheim is commonly taken to appeal to the causal overriding criterion when maintaining that social phenomena exist *sui generis* (see Durkheim 1897). The opposite view is particularly associated with Popper's and Watkins' work from around the 1950s. Both Popper and Watkins saw the claim that social phenomena have causal powers that override the powers of individuals as going together with the historicist view that societies go through stages of development according to deterministic laws of historical development, that is, laws which individuals cannot change in any way (see Popper 1964 [1957] and Watkins 1973 [1957]). Popper argued that there cannot be such laws and as such he rejected the idea of the *sui generis* existence of social phenomena relative to the causal overriding criterion. Turning to the translation criterion, it was appealed to in a much discussed paper by Mandelbaum from 1955 (Mandelbaum 1973 [1955]). Mandelbaum maintained that it is impossible to translate terms which refer to social phenomena into terms which refer to the thoughts and actions of individuals only. This shows, he contended, that social phenomena are not identical with specific individuals having certain thoughts and performing given actions. Hence, he concluded that social phenomena should be regarded as existing *sui generis*.

Nowadays there are few, if any, theorists who defend the theses of ontological holism or individualism by appeal to the causal overriding and the translation criteria. Instead, the majority of theorists tend to be concerned with the composition and determination criteria. More specifically, most participants in the debate regard themselves as ontological individualists by reference to one or both of these criteria: They subscribe to the view that social phenomena are composed of and/or non-causally determined by individuals. But what exactly does it mean to say that social phenomena are non-causally determined by individuals? This question has received a lot of attention in the recent debate. One influential suggestion is to spell out this relation as one of supervenience. Applied to social phenomena, this is the contention that there can be no difference at the level of social phenomena without there being a difference at the level of individuals. Roughly speaking, this means that if individuals with certain properties, standing in certain relations, etc. instantiate a university, then they cannot cease to do so unless some change takes place as to their properties, interrelations, and so on. This way of spelling out the notion of supervenience is only one among many. Moreover, various other notions, such as the notions of emergence, realization, and grounding, have been brought into play in the attempt to cash out the notion of non-causal determination. Their exact specification, too, is a matter of discussion.

The wide acceptance of the view that social phenomena are non-causally determined by individuals means that the most recent criteria of *sui generis* existence are made in full compatibility with this point. To illustrate this observation, consider the agency criterion as currently defended by Christian List and Pettit (see, e.g. List and Pettit 2011). The thought behind the criterion is that in order for social organizations and the like to qualify as group agents, it must be possible to ascribe to them certain attitudes supervenient upon the attitudes of individuals. It is very difficult, though, to pin down the exact individual attitudes that these attitudes of group agents supervene

upon. For this reason, List and Pettit contend, social phenomena that are group agents should be regarded as existing over and above individuals.

In discussions of the ontological status of social phenomena, the term “explanation” is sometimes used to refer to ontological analyses of how individuals must be related, what properties they must have, and so on, in order to constitute social phenomena of various sorts. Explanations of this sort map out non-causal or synchronic relationships between social phenomena and individuals. They should be distinguished from explanations which map out the causal or diachronic relations between events or states involving social phenomena and individuals. It is explanations in the causal or diachronic sense that is the focus of the methodological individualism-holism debate.

## 1.2 The Methodological Individualism-Holism Debate

Why is the European economy in recession? Why has the birth rate in Tanzania recently gone up? And why do revolutions tend to be followed by famines? The methodological individualism-holism debate is about the proper focus of social scientific explanations that are advanced in response to questions like these. More specifically, it concerns the extent to which social scientific explanations may, and should, focus on individuals and social phenomena respectively. The discussion takes various forms. Among these, there are in particular two which stand out: The dispensability debate, as it may be called, and the microfoundations debate.

### 1.2.1 The Dispensability Debate

The debate about dispensability revolves around the question of whether individualist (or individual level) or holist (or social level) explanations may, and should, be dispensed with within the social sciences. There are three basic positions on this issue:

*Methodological individualism:* Individualist explanations should be advanced. Holist explanations may, and should, be dispensed with.

*Strong methodological holism:* Holists explanations should be offered. Individualist explanations may, and should, be dispensed with.

*Weak methodological holism:* Not only individualist but also holist explanations should be put forward. Neither individualist nor holist explanations may, and should be, dispensed with.

In order to further characterize the methodological debate, it is instructive to go over four dimensions along which particular methodological individualist and holist positions may vary. To begin with, particular positions may differ with respect to their notion of explanation. All theses of methodological individualism

and holism express a view as to what is the proper focus of explanation. By implication, they involve some view of what an explanation is. In line with the general trend in philosophy of science, there was a point at which most theorists endorsed the Covering Law model of explanation. According to this model, a scientific explanation takes the form of a deductive or inductive argument which shows why an event in need of explanation was to be expected. Today, this view is rarely, if ever, espoused. Instead, a number of different notions of explanation are appealed to in the debate. For instance, some participants in the debate embrace the erotetic model of explanation which states that an explanation is an answer to a why-question. Others subscribe to the causal-information view of explanation which has it that an explanation provides information about the causal process leading to the event in need of explanation, and so on.

Whatever their notion of explanation, participants in the methodological debate make a distinction between individualist and holist explanations. It is widely concurred that individualist explanations revolve around individuals, their beliefs, actions, etc. whereas holist explanations are centered round social phenomena such as social organizations, structures, and the like. Beyond this point, however, agreement ceases: There is no general consensus as to how, more precisely, to circumscribe individualist and holist explanations. Here are some of the issues that have been discussed: What kinds of relations and interactions among individuals should individualist explanations be allowed to refer to? Does an explanation really qualify as holist if it describes the rules and norms within a group of individuals? Are individualist explanations only permitted to refer to particular individuals or may they describe types of individuals too? The debate on the proper distinction between individualist and holist explanations has primarily concentrated on the category of individualist explanations. One reason for this is probably that as the debate evolved, methodological individualists have opted for ever broader conceptions of individualist explanations. As such, they have continuously brought up the question of the proper circumscription of these explanations.

Both individualist and holist explanations may be divided into different types. And not surprisingly, there are various conceptions as to what constitutes an adequate or satisfactory type of individualist or holist explanation. For example, one type of individualist explanation that has been highly popular among methodological individualists is explanations informed by rational choice models. These state how rational individuals acted in light of their beliefs and desires. Another type of individualist explanation consists in accounts of individuals' actions by appeal to their dispositions to act in certain ways in certain circumstances. Turning to holist explanations, the advancement of accounts that point to the statistical properties of groups has been very prevalent among methodological holists as have explanations specifying how social organizations bring about various effects. Needless to say, there are multiple more or less fine-grained ways in which to differentiate between types of individualist and holist explanations. It should be stressed that the rejection of the adequacy of, say, explanations informed by rational choice theory or explanations by appeal to statistical properties does not amount to a refutation of methodological individualism or holism as such. All it amounts to

is the repudiation of the adequacy of a particular type of individualist or holist explanation.

Conceptions of adequate individualist and/or holist explanations typically go together with a preference for the use of one or several forms of explanations. Three commonly mentioned forms of explanations are functional explanations, intentional explanations, and straightforward causal explanations, as they may be called. In the past, methodological holists have often favored the use of functional explanations. Accordingly, they have explained the continued existence of various social phenomena by reference to their function, or effect, in some society. By contrast, methodological individualists frequently make use of intentional explanations: They explain actions by appeal to individuals' reasons or motivations for their actions. Today, functional and intentional explanations are often regarded as a species of causal explanations while being contrasted with straightforward causal explanations. When applied by methodological holists, straightforward causal explanations state, say, that a rise in unemployment caused a rise in crime. Again, different and more fine-grained classifications of forms of explanation have also been suggested.

As these considerations bring out, there is a variety of ways in which particular individualist and holist positions may differ. It is possible to distinguish between particular individualist and holist positions as regards their notion of explanation, their distinction between individualist and holist explanations, their conception of what constitutes adequate types of individualist and holist explanation, and their preference when it comes to different forms of explanation.

Turning now to a description of the dispute itself, there are few proponents of strong methodological holism. As a result, the debate mainly plays itself out between methodological individualists and weak methodological holists. As both parties agree that individualist explanations should be advanced, methodological individualists put all their efforts into showing that holist explanations may, and should, be dispensed with, whereas weak methodological holists concentrate on showing that holist explanations may not, and should not, be dispensed with. A number of arguments have been offered in support of, and against, these stances.

It is worth registering a few points relating to the development of the debate. Sometimes methodological holism has been linked with the espousal of collectivist or totalitarian political ideals. For example, Hayek and Popper drew this connection when advancing their views around the 1950s. They saw liberalism as based upon a commitment to methodological individualism and collectivism or totalitarianism as being underpinned by a commitment to methodological holism. Today, it is widely held that there is no necessary linkage between being either a methodological individualist or holist, and having a certain political orientation. Thus, discussions of methodological individualism and holism typically take place without any reference to political values of any form.

Another point to notice is that a particular way of approaching the defense of both methodological individualism and methodological holism has been quite dominant within the debate. In 1961, Ernest Nagel published *The Structure of Science* that contains his influential model of intertheoretic reduction. In the period

following its publication, theorists engaged in the methodological individualism-holism debate became increasingly concerned with reduction. Many of them began to regard the question of whether holist explanations are dispensable as a matter of whether holist theories are reducible to individualist ones in accordance with Nagel's model of intertheoretic reduction.

In order to apply this model to the methodological individualism-holism debate, the following assumption is made: Holist explanations involve holist theories which are distinguished by their use of holist terms such as "university" or "school." By contrast, individualist explanations involve individualist theories which contain descriptions of individuals only. Reduction is conceived of as a two-step procedure. First, the holist terms must be linked, on a one-to-one basis, with descriptions of individuals via bridge laws. The bridge laws express that the linked predicates are co-extensive, i.e. have the same reference, in a lawlike manner. Next, the holist theory must be deduced from, and in this sense explained by, an individualist theory plus the bridge laws. If both these conditions are met, the holist theory has been reduced to an individualist theory and holist explanations that refer to the holist theory may be dispensed with: It is possible to replace these explanations with individualist explanations that involve the reducing individualist theory. By contrast, if the holist theory is not reducible, then explanations invoking it may not be substituted by individualist explanations that appeal to individualist theories. When engaged in this dispute, methodological individualists think that all holist theories are reducible to individualist ones whereas weak methodological holists deny this.

Discussions of the possibility of reduction were probably at their peak in the 1980s and 1990s. In addition to Nagel's original model of intertheoretic reduction, a number of modified versions have been proposed. All conceptions of reduction in the Nagel family, however, have been seriously disputed to the point where many theorists today consider any interest in the possibility of intertheoretic reduction as misguided. This means that in the current debate, the question of whether holist explanations are dispensable tends to a much lesser extent to be phrased as a matter of the possibility of intertheoretic reduction.

Finally, it may be observed that the dispensability debate from the 1980s and onwards is marked by the appearance of a number of new, or new versions of, arguments in support of weak methodological holism. Most notably, these include the defense of the intertheoretic irreducibility of holist theories and, by implication, holist explanations by reference to the argument from multiple realization (see Kincaid 1996, 1997); the defense of the indispensability of holist explanations by appeal to the notion of emergence (see Bhaskar 2000 [1979]); and the argument that holist explanations are indispensable because they satisfy theoretical and practical interests that cannot be satisfied by corresponding individualist explanations (see Garfinkel (1981) and Jackson and Pettit (1992)). The advancement of these arguments has not resulted in methodological individualism no longer being endorsed and defended. Still, they have contributed to the situation today in which methodological individualism is no longer the position that obviously dominates the field.

### 1.2.2 *The Microfoundations Debate*

The microfoundations debate deals with the question of whether holist causal claims need microfoundations: Should holist causal claims be supplemented by accounts of underlying mechanisms at the level of individuals? There are two basic positions on this issue:

*Methodological individualism:* Holist causal claims should be supplemented by accounts of the underlying mechanisms at the level of individuals.

*Methodological holism:* Holist causal claims need not be supplemented by accounts of the underlying mechanisms at the level of individuals.

To gain a better understanding of the dispute, it is instructive to begin by saying a bit about the notion of holist causal claims. Holist causal claims are ones in which both the cause and the effect are described in holist terms. In turn, holist terms are often specified as descriptions of the statistical properties of a group such as “a high literacy rate” or “a low suicide rate.” Also, they are sometimes taken to be exemplified by, e.g., terms like “university” or “prison” that refer to social organizations. Moreover, note that holist causal claims should be taken broadly to include not only straightforward causal claims, as they may be called. These are illustrated by the assertion that the high rate of unemployment caused a high crime rate. Also, and most notably, they should be taken to comprise what is commonly regarded as a special kind of causal claims, namely functional claims. An example of a functional claim is the contention that the state continues to exist because it has the function, or effect, of furthering the interests of the ruling classes. All varieties of holist causal claims, methodological individualists assert, should be supplemented by accounts of the underlying mechanisms at the level of individuals.

So what then are individual level accounts? They are descriptions of the chain of events, at the level of individuals, that link the cause and effect described in holist terms. Sometimes, it is further maintained that these accounts must specify the laws or law-like regularities that govern the transitions between these events. There has been some, but not much, discussion of what exactly to regard as accounts at the level of individuals: Most participants in the debate work with rather inclusive notions of how individuals may be described. Likewise, it may be registered that particularly one view on what constitutes adequate individual level accounts has been prominent, namely formulations of individual level accounts based on rational choice models.

There are different ways in which to qualify the claim about the need for supplementary individual level accounts. For instance, some methodological individualists insist that holist causal claims should be accompanied by individual level accounts in order to count as explanations. These theorists subscribe to the mechanism view of explanation which states that to explain is to provide an account of the underlying mechanisms linking a cause and its effect. Other methodological individualists insist only that holist causal claims do not qualify as *complete* explanations unless individual level accounts are supplied. And yet others confine themselves to



asserting that holist explanations are *less satisfying* if they are not complemented by individual level accounts. However qualified, methodological holists hold that the thesis of methodological individualism is wrong. A number of arguments have been advanced both in support of, and against, these various positions.

The microfoundations debate gained prominence in the 1980s when a number of philosophers and social scientists began to defend the need for microfoundations as a new form of methodological individualism. Important contributions to this debate include Boudon (1998), Coleman (1986), Elster (1985), and Little (1986). Today, a number of social theoretic approaches, such as analytical sociology and public choice theory, insist on the need for microfoundations. The debate is still well and alive.

### 1.3 Two Related Debates

By way of ending this overview, the existence of two related debates may be briefly noticed: The micro-macro debate and the agency-structure debate.

The micro-macro debate revolves around a distinction between micro and macro level phenomena, whereas the agency-structure debate works with a differentiation between agents and structures. Depending on how these distinctions are drawn, they may line up with particular distinctions between individuals and social phenomena. Further, the micro-macro debate involves a distinction between micro and macro level explanations, just as the agency-structure debate makes a distinction between explanations that focus on agents and structures respectively. Again, depending on how these distinctions are specified, they may map onto particular distinctions between individualist and holist explanations. Finally, all three debates address many of the same basic ontological and methodological issues.

This said, it is also important to draw attention to some of the differences between the debates. As already indicated, the above distinctions pertaining to ontology and explanation are far from always drawn in the same way. Also, there are many ontological and methodological issues that are not discussed, or at least not to the same degree, within all three debates. For example, the agency-structure debate has significantly focused on the issue of whether – and how – social structures restrict, or even compromise, individual agents' autonomy. This question has received comparatively little attention within both the micro-macro and the recent individualism-holism debate. Or, to mention one last example, the individualism-holism debate is rather alone in having paid so much heed to the possibility of intertheoretic reduction. In light of these considerations, the individualism-holism, the micro-macro, and the agency-structure debates may be characterized as three more or less distinct debates with significant overlaps.

## 1.4 The Book in Outline

We now turn to the task of briefly presenting the papers in this collection. Before doing so, however, two comments are in order. One concerns the categorization of the papers. We have divided them according to whether they are *primarily* contributions to the ontological or the methodological individualism-holism debate. Of course, many papers make points pertaining to both these debates. The other comment is about terminology. There is no consensus in the individualism-holism debate about the use and definition of many of the key terms. In the preceding overview, we have suggested one way of labeling and defining key positions and views. However, there are alternative ways. This is worth keeping in mind when reading the following outline of the papers and the papers themselves.

There are six papers which are chiefly concerned with the ontological version of the individualism-holism debate. The first five of these reflect on, and discuss, the ontological status of social phenomena with a special focus on their relationship to individuals.

In “*What Is Individualism in Social Ontology? Ontological Individualism vs. Anchor Individualism*,” Brian Epstein considers two theses: Ontological individualism specified mainly as the claim that social facts supervene upon individualistic facts, and anchor individualism specified as the view that social facts are exhaustively anchored by individualistic facts. So far this latter thesis has received surprisingly little attention in the ontological debate. Epstein systematically discusses and compares the two theses while also pointing to a tension between them.

Next follows Dave Elder-Vass’ paper “*Social Entities and the Basis of their Powers*” centered round the widely discussed notion of emergence. Elder-Vass defends a relational theory of emergence and shows how this theory – despite frequent claims to the contrary – makes it possible to justify the ascription of causal powers to social entities. Elder-Vass points out that this conclusion amounts to a rejection of methodological individualism specified as the view that structural causation is impossible. Also, he shows how his theory of emergence makes it possible to think in new ways about social structures.

Daniel Little’s “*Actor-Centered Sociology and the New Pragmatism*” takes as its starting point the observation that social phenomena are constituted by the actions and thoughts of socially constituted and socially situated individuals. Little contends that an adequate theory of the actor is needed in order to get a better grasp of how individuals constitute social phenomena and he shows how such a theory may be extracted from the work by current pragmatist sociology. Also, he points to a number of reasons why social scientists should look into the manner in which individuals make up social phenomena.

“*Three Doctrines in Social Ontology*” by Philip Pettit is devoted to the defense of three ontological theses. These are: Individualism – the view that social wholes do not operate under laws that compromise individuals’ intentional psychologies; anti-atomism – the claim that there are aspects of our intentional psychology that depends noncausally on social interaction; and anti-singularism – the view that

social wholes may qualify as (group) agents with attitudes that supervene on the attitudes of their members.

Like Pettit, Andras Szigeti is concerned with group agents in his paper “*Collective Responsibility and Group Control.*” Szigeti examines the claim that group agents or collective agents, as he calls them, are sometimes robustly responsible, i.e. their moral responsibility is something over and above the combined moral responsibility of their members. Plausibly, collective agents can only be held responsible for something that they have some control over. Yet, Szigeti argues, the idea that collectives can be in control is problematic whether collective control is understood as noncausal group control or as causal group control.

The sixth and last paper in the section on ontology takes issue with an assumption informing much of the debate, namely that the distinction between the micro and macro should be conceived of as a fixed and categorical distinction between two levels. In his paper, “*Rethinking Micro-Macro Relations,*” Petri Ylikoski argues that this conception should be replaced by a context-relative and noncategorical distinction between the micro and macro as a difference of scale. Ylikoski shows how there are numerous gains associated with the adoption of this new approach to the micro-macro distinction.

The rest of the papers in this book primarily address different aspects of the methodological individualism-holism debate. Four of the papers in this category involve discussions of what constitutes the right level of explanation.

In his paper “*Dead Ends and Live Issues in the Individualism-Holism Debate,*” Harold Kincaid argues that we should pursue the issues discussed within the individualism-holism debate as they arise in the context of actual social scientific research. He shows how the failure to follow this strategy leads to dead ends. Also, he illustrates this approach, among other things, by examining social scientific debates where the focus is not whether to explain purely in terms of individuals or purely in terms of social structures, but rather the extent to which facts about social structure must be added to facts about individuals in order to successfully explain social phenomena.

Jeroen van Bouwel’s “*Explanatory Strategies Beyond the Individualism/Holism Debate*” likewise observes that social scientists make use of a plurality of explanatory approaches. By use of the erotetic model of explanation, Van Bouwel shows how different forms of explanation at different levels are indispensable if we want an answer to all our explanation seeking questions. Accordingly, he insists, it is time that methodological individualists and holists give up discussions as to who is right and concentrate instead on the questions of how to conceive of the relationship between explanations at different levels and when to use which explanatory strategy.

In turn, Julie Zahle’s contribution, “*Holism, Emergence and the Crucial Distinction,*” contains a discussion of Elder-Vass’s argument from emergence in support of holist explanations being indispensable. Zahle argues that current methodological individualists should reject the argument because it presupposes a distinction between individualist and holist explanation that they find unacceptable. Zahle continues by urging that participants in the debate should begin to justify

their particular distinction between individualist and holist explanations on pragmatic grounds.

Finally, Finn Collin's paper "*Who Are the Agents? Actor Network Theory, Methodological Individualism, and Reduction*" examines Latour's Actor Network Theory in the perspective of the individualism-holism debate. Collin traces the development in Latour's thought from his holist emphasis on macro explanations over an individualist concern with explanations focusing on individuals to his current position centered round the notion of an actant. Collin links these developments to broader trends within the individualism-holism debate.

The last two papers in this section deal with the question of what constitutes a satisfactory notion of individual agency when offering explanations or accounts at the level of individuals.

Mark Risjord's paper "*Structure, Agency, and Improvisation*" takes as its starting point an issue which has been intensively discussed within the agency-structure debate: How do individuals create and reproduce structures and how do structures influence individuals? Risjord shows how the explanations in response to these questions as offered by microfoundationalist approaches and practice theories rely on inadequate notions of individual agency. In their place, he offers an account of agency as ecological attunement that incorporates elements from both traditions.

Finally, Matthew McCubbins and Mark Turner's paper "*Are Individuals Fickle-Minded?*" shows how game theorists, including behavioral game theorists, are methodological individualists who assume certain kinds of consistency in individual actors. Without this assumption of consistency, game theory does not have any predictive power. By appeal to a number of experiments, the authors argue that this presupposition is unwarranted. They conclude by recommending that more realistic models of individual actors must be developed that acknowledge the variance in behavior for a given individual.<sup>3</sup>

## References

- Bhaskar, R. (2000 [1979]). *The possibility of naturalism*. London: Routledge.
- Boudon, R. (1998). Limitations of rational choice theory. *American Journal of Sociology*, 104(3), 817–828.
- Coleman, J. S. (1986). Social theory, social research, and a theory of action. *The American Journal of Sociology*, 91(6), 1309–1355.
- Durkheim, E. (1897). *Le suicide, etude de sociologie*. Paris: Félix Alcan.
- Elster, J. (1985). *Making sense of Marx*. Cambridge: Cambridge University Press.
- Garfinkel, A. (1981). *Forms of explanation*. New Haven: Yale University Press.

---

<sup>3</sup> All sections, except Sect. 1.4, have been written by Julie Zahle with Finn Collin providing helpful suggestions. Section 1.4 has been co-written. Also, many thanks to Harold Kincaid and Daniel Little for their useful comments on this introduction. Julie Zahle's work on the paper was supported by The Danish Council for Independent Research | Humanities (FKK).

- Jackson, F., & Pettit, P. (1992). In defense of explanatory ecumenism. *Economics and Philosophy*, 8, 1–21.
- Kincaid, H. (1996). *Philosophical foundations of the social sciences*. New York: Cambridge University Press.
- Kincaid, H. (1997). *Individualism and the unity of science*. Lanham: Rowman & Littlefield.
- List, C., & Pettit, P. (2011). *Group agency*. Oxford: Oxford University Press.
- Little, D. (1986). *The scientific Marx*. Minneapolis: University of Minnesota Press.
- Mandelbaum, M. (1973 [1955]). Societal facts. In J. O'Neill (Ed.), *Modes of individualism and collectivism* (pp. 221–234). London: Heineman.
- Popper, K. R. (1964 [1957]). *The poverty of historicism*. New York: Harper & Row.
- Udehn, L. (2002). The changing face of methodological individualism. *Annual Review of Sociology*, 28, 479–507.
- Watkins, J. W. N. (1973 [1957]). Historical explanation in the social sciences. In J. O'Neill (Ed.), *Modes of individualism and collectivism* (pp. 166–178). London: Heineman.
- Zahle, J. (2007). Holism and supervenience. In S. Turner & M. Risjord (Eds.), *Handbook of philosophy of anthropology and sociology* (pp. 311–341). Amsterdam: Elsevier.
- Zahle, J. (2013). Holism in social sciences. In B. Kaldis (Ed.), *Encyclopedia of philosophy and the social sciences*. London: Sage Publications.

**Part I**  
**Ontological Individualism-Holism**

## Chapter 2

# What Is Individualism in Social Ontology? Ontological Individualism vs. Anchor Individualism

Brian Epstein

**Abstract** Individualists about social ontology hold that social facts are “built out of” facts about individuals. In this paper, I argue that there are two distinct kinds of individualism about social ontology, two different ways individual people might be the metaphysical “builders” of the social world. The familiar kind is *ontological individualism*. This is the thesis that social facts supervene on, or are exhaustively grounded by, facts about individual people. What I call *anchor individualism* is the alternative thesis that facts about individuals put in place the conditions for a social entity to exist, or the conditions for something to have a social property. Examples include conventionalist theories of the social world, such as David Hume’s theories of promises, money, and government, and collective acceptance theories, such as John Searle’s theory of institutional facts. Anchor individualism is often conflated with ontological individualism. But in fact, the two theses are in tension with one another: if one of these kinds of individualism is true, then the other is very unlikely to be. My aim in this paper is to clarify both, and argue that they should be sharply distinguished from one another.

## 2.1 Introduction

It is widely recognized that individualism—a topic long debated in the social sciences—consists of two different theses. *Methodological (or explanatory) individualism* is a thesis about the methodology of the social sciences: it holds that explanations of social facts or phenomena should be individualistic. That is, they should be given in terms of individual people and certain relations between

---

B. Epstein (✉)

Department of Philosophy, Tufts University, Medford, MA, USA

e-mail: [brian.epstein@tufts.edu](mailto:brian.epstein@tufts.edu)

individuals. *Ontological individualism*, on the other hand, is a thesis about the nature or metaphysics of social facts or phenomena. It holds that social facts or phenomena are exhaustively built out of, or depend on, individualistic ones. To put it in the familiar way, ontological individualism is the thesis that there is nothing to social facts “over and above” facts about individuals and certain relations between individuals. In recent decades, these two theses have been thoroughly disentangled from one another. Many theorists endorse methodological individualism, while many others deny it. But nearly everyone agrees with ontological individualism, believing it to be the only alternative to the crazy idea that social facts are autonomous spirits, lurking in a separate metaphysical realm.

In recent work, I have challenged ontological individualism. The thesis is not as trivial as many people assume, nor is it the only alternative to a crazy metaphysics of the social. In fact, I argue in Epstein 2009 that even on a most charitable interpretation, the thesis is false.

There are, however, two quite different ways the social world might “depend on” or be “built out of” facts about people. The literature on ontological individualism—my 2009 article included—only notices one of them. My aim in this piece is to distinguish ontological individualism from a different thesis about the role of individuals in “building” the social world. This second thesis I will call *anchor individualism*. Both are metaphysical theses about the relation of social facts to facts about individual people.

Ontological individualism is a thesis about the *composition*, *supervenience*, *exhaustive determination*, or *grounding* of social facts by individualistic ones. As Harold Kincaid puts it:

Social wholes are both composed of individuals and determined by their actions. . . . Individuals determine the social world in the intuitive sense that once all the relevant facts (expressed in the preferred individualist vocabulary) about individuals are set, then so too are all the facts about social entities, events, etc. Or, to put this idea in terms of supervenient properties, the social supervenes on the individual in the sense that any two social domains exactly alike in terms of the individuals and individual relations composing them would share the same social properties. (Kincaid 1986, p. 499)

In short, ontological individualism is the thesis that social facts supervene on or are exhaustively grounded by individualistic ones. A different way of understanding the metaphysical relation between individuals and the social world, however, does not take social facts to be composed of or supervenient on the individual. Rather, it derives from the idea that things have social properties in virtue of people conceiving of them in a certain way, or treating them in a certain way, or acting in a certain way. This alternative thesis is that facts about individuals put in place the conditions for a social entity to exist, or the conditions for something to have a social property. Although this alternative is less commonly discussed, it is at the heart of many contemporary approaches to the ontology of the social world, and it has deep historical roots. To coin a new term, I will call this “putting in place” relation ‘anchoring’, a term I will define more carefully in Sect. 2.3. *Anchor individualism*, then, is the thesis that social facts are exhaustively anchored by individualistic ones. Facts about individuals and certain relations between them, the



thesis holds, set up the conventions, do the projecting, or otherwise anchor the social world.

Although ontological individualism and anchor individualism are often conflated, they are independent of one another: it is coherent to endorse one while denying the other. That is not to say that they are unrelated. In fact, they are in tension with one another. If one of these kinds of individualism is true, then the other is very unlikely to be. My aim in this paper is to describe the two theses and identify the elements that must be clarified if the theses are to be pinned down, propose that the theses are distinct, and point out the tension between them. In this paper, I will not address the truth or falsity of either thesis.

## 2.2 Ontological Individualism

Ontological individualism is a thesis in the field known as “inter-level metaphysics.” Inter-level metaphysics is concerned with the relations between objects, events, properties, and facts at different “levels.” Some of the problems of inter-level metaphysics have to do with *causation* at different levels, such as whether it is coherent to take a high-level event as a cause for a low-level one,<sup>1</sup> as in the proposition, *Jane’s belief that there was a fly on her nose caused such-and-such a neuron to fire.*

A different set of problems of inter-level metaphysics has to do with *non-causal* determination. Jane’s belief that there was a fly on her nose had certain causal consequences. But the belief itself is arguably “built out of” her neural states being in a certain pattern. A large family of terms is used to describe different ways that one entity may be “built out of” another: the belief *is composed of* neural states, the belief *is constituted by* neural states, the neural states *realize* the belief, the belief *emerges from* the neural states, the belief *is grounded by* the neural states, or the belief *supervenes on* neural states (Bennett 2011, p. 81). Part of the task of clarifying ontological individualism is deciding which of these “building” relations are the pertinent ones.

To argue that high-level entities like belief are “built out of” low-level entities like neural states, there are several projects we might pursue. One is to consider them token-by-token: what particular pattern of neural states grounds Jane’s particular belief about the fly? Another is to consider them type-by-type: what type of pattern of neural states in a person’s mind grounds a belief in that person that she has a fly on her nose? Both of these, however, are often difficult or impossible to answer, and anyway involve more detail than we need. A more straightforward project is to consider the relation between the high-level and low-level domains as a whole, or between sets of facts or properties at one level and sets of facts or

---

<sup>1</sup> These problems arise when we assume—plausibly enough—causal closure of the physical and a lack of widespread causal overdetermination. See Kim 1998, p. 2.

properties at another. For instance, we might just argue that *all* of a person's mental states in general—including beliefs, desires, and so on—are “built out of” that person's neural states. This is a generic thesis about the determination of the whole set of mental states, somehow or other, out of a whole set of neural states, without trying to match a particular type of belief or desire with a particular type of neural state.

This generic strategy is what social theorists employ in advocating ontological individualism. Ontological individualism does not require that we give the individualistic properties that correspond to each social property, like *being an investment bank*, *being President of the U.S.*, or *having cashed a check*. Instead, ontological individualism is the general thesis that the social world is “built” out of an “individualistic base,” or that social facts in general are built out of facts about individual people.

That is a very rough way of putting the thesis, leaving many things vague. What, for instance, counts as the “individualistic base,” or the “facts about individual people”? Do these include people's mental properties, their physical properties, and/or certain relations between them? Different versions of the thesis of ontological individualism will take positions on this. Some theorists hold that social facts are built out of psychological states of individuals. Others hold that they are built out of both psychological states and certain relational properties between them. Still others hold that they are built out of behaviors, activities, or practices.

Similarly, we might ask which “building” relation is claimed to hold, between the social world and the individualistic base. Are social things *composed out of* individual people? Or instead, are social facts *grounded by* individualistic facts? And further, we might ask which “social facts” the thesis is about. Institutions? Artifacts? Groups?

In light of these ambiguities, it is more accurate to think of ontological individualism as a family of theses rather than a single one, or else as a very vague thesis that needs to be made precise in order to be understood and assessed. Ontological individualism is a thesis about non-causal determination in inter-level metaphysics, in particular, about some “exhaustive building” relation between individual-level stuff and social-level stuff. To be more precise than this, however, we need to say more about these three topics:

1. What the “exhaustive building” relation is, which is claimed to hold between the individualistic base and the social level;
2. What the individualistic base is: individual people, their mental states, their actions, habits, or practices, various relations among them, etc.; and,
3. What is included in the social level: the facts, events, objects, and properties held to be “built” out of the base.

I will discuss each of these in turn, with the aim of raising the key issues, not resolving a single correct response for each.

### 2.2.1 *The “Exhaustive Building” Relation*

Many theorists, in discussing ontological individualism, speak of what the social world “consists of.” For instance, in a well-known 1968 paper Steven Lukes says, “Society consists of people. Groups consist of people. Institutions consist of people plus roles and rules” (Lukes 1968, p. 451). Geoffrey Hodgson, in a 2007 paper, uses the same term to make a different claim: “For reasons that will become clearer below, it is crucial whether it is claimed that the social world simply consists of individuals, or of individuals *and interactive relations between them*” (Hodgson 2007, p. 215). In the next section I will discuss things like interactive relations between individuals, but let us begin by noticing that we have to do better than “consists of.”

There are many senses in which one thing might “consist of” another. A promissory note might consist of its material constitution, e.g., a blob of paper and ink. Or it might consist of its parts, or it might consist of its wording, or it might consist of the promise with which it was issued, or it might consist of its essential properties, and so on. Another problem with “consists of” is that it seems only to apply to objects, not properties or facts. It may not even be grammatical to speak of a social property like *being President* or a social fact like *Obama is President* “consisting of” properties of individuals or facts about individuals. We need a more precise notion.

The most sophisticated treatments of ontological individualism, like Harold Kincaid’s I quoted above, have latched onto *supervenience*. Supervenience is a family of relations, designed expressly for clarifying non-causal determination between sets of entities at one level and sets of entities at another. Usually, supervenience is understood as a relation between two sets of properties—chemical properties and physical properties, mental properties and neural properties, social properties and individualistic properties, and so on. However, it can relatively easily be adapted to relate sets of facts, events, or objects, instead of properties.

There are many forms of supervenience, but the most pertinent to ontological individualism is *global supervenience*.<sup>2</sup> Consider two sets of properties, a high-level set of properties A, and a low-level set of properties B. Intuitively, to say that A globally supervenes on B is to say that if we fix all of the B-properties in the whole world, then that suffices to fix all the A-properties in the world as well. To say that the chemical properties globally supervene on the properties of microphysics is to say that if it is fixed what all the microphysical properties are, everywhere in the world, then there is no more work to do in fixing the chemical properties. They are already fixed.

To be more precise still, the notion of global supervenience is cashed out in terms of possible worlds. A common formulation is:

---

<sup>2</sup>Currie 1984 points out that local supervenience fails, for social properties. See Epstein 2008, 2011 for refinement of this point.

A-properties globally supervene on B-properties if and only if for any worlds  $w_1$  and  $w_2$ , if  $w_1$  and  $w_2$  have exactly the same worldwide pattern of distribution of B-properties, then they have exactly the same world-wide pattern of distribution of A-properties.<sup>3</sup>

Applying the notion to ontological individualism, suppose we have come to resolution on what the social properties are and what the individualistic properties are. To say that the social properties globally supervene on the individualistic properties is to say that once the worldwide distribution of individualistic properties, present *and* historical, is fixed, that already fixes the worldwide distribution of social properties. This is an intuitive way to capture the idea that there is nothing to the social “over and above” the individualistic.

Global supervenience is a very minimal relation. That is, if it does hold between two sets, there does not need to be any way of identifying which individualistic properties determine which social properties in which worlds. And it certainly does not imply anything about individualistic explanation of social phenomena. More technically, supervenience is also only a claim about co-variance between worlds—that is, if one world has a *different* worldwide distribution pattern of social properties than another, then the two worlds must have different worldwide distribution patterns of individualistic properties.

This minimal-ness is both an advantage and disadvantage to global supervenience, as a way of understanding the “exhaustive building” relation in ontological individualism. The advantage is that it sets the bar nice and low. If global supervenience does hold, it provides at least some sense in which the social properties depend on the individualistic ones, without requiring that each social property correspond to particular individualistic ones. The disadvantage, however, is that the bar is so low, that it can barely be counted as saying that the social world is *determined by* the individualistic, or is *built out of* the individualistic. Supervenience is only a claim about co-variances: changes in the social must be accompanied by changes in the individualistic. Many philosophers have recently argued that we need more. Just because two sets of properties co-vary in the right way does not suffice to show that one is built out of the other. Thus it is not quite right to understand the “exhaustive building” relation as supervenience. Instead, all the different supervenience relations are best understood as diagnostic tools, like X-ray or MRI machines, for assessing when an exhaustive building relation is in place.

Increasingly prominent among metaphysicians is a new approach to the basic “building” relation: what is being called the *grounding* relation (Fine 2001; Rosen 2010; Schaffer 2009). The fact *Obama is President*, for instance, is grounded by a set of other facts, such as the fact that he is at least 35 years old, that he won a majority of votes by the Electoral College, and that he took the appropriate oath of office. There are other facts that his being President might co-vary with, but this set of grounding facts is the metaphysical basis for the fact that he is President.

---

<sup>3</sup> McLaughlin and Bennett 2005. There are, in fact, several versions of global supervenience, but their differences are not important for our purposes; cf. Epstein 2009.

The literature on grounding is reasonably new. It is promising, because it has the potential to help clarify when one fact is the case *in virtue of* another fact being the case. But philosophers are still working out the details of the idea. Moreover, grounding alone is not enough to be the “exhaustive building” relation we need. The grounding relation holds between some fact *F* and a set of facts {*G*<sub>1</sub>, *G*<sub>2</sub>, . . .} that together ground *F*. But remember that ontological individualism is a generic claim about the relation of the whole set of social facts to the whole set of individualistic facts. That is what global supervenience is trying to capture—this generic dependence of a whole set of facts on another whole set of facts. Thus the grounding relation should be seen as a replacement for the “covariance” relation that is part of a supervenience claim. To capture the idea that the entire set of social facts is “exhaustively grounded” by individualistic facts, we will need to construct something like the supervenience relation, but using grounding as a building-block rather than covariance.<sup>4</sup>

All that is to say that for the moment, global supervenience is a reasonable way to understand the “exhaustive building” relation in ontological individualism. But further work will likely permit global supervenience to be replaced with an improved understanding.

### 2.2.2 *The Individualistic Base*

A more difficult problem is pinning down what the individualistic base is, in a claim of ontological individualism. Consider again the quotation from Hodgson: “it is crucial whether it is claimed that the social world simply consists of individuals, or of individuals *and interactive relations between them.*” As I discussed in the last section, we would do better to replace ‘consists of’ with ‘globally supervenes on’. Putting it in these terms, Hodgson can be read as suggesting that there is a crucial difference between the claims:

- (a) The social world globally supervenes on individuals, and
- (b) The social world globally supervenes on individuals and interactive relations between them.

This is a claim about the “individualistic base” for the social world. When the ontological individualist says that the social supervenes on the individualistic, how should ‘individualistic’ be understood?

Hodgson’s argument is useful to consider, because it is a widely held perspective on individualism. His basic line is this. If we only consider individuals in isolation, we do not have the complete building blocks for the social, since that would neglect their interactions. Furthermore, when we talk about “interactive relations,” we need to think about two different things: the individual interactions between pairs of

---

<sup>4</sup>Correia 2005 develops a number of candidates.

people, and also the social relations that people stand in. For instance, if person X purchases something from you, X interacts one-to-one with you, but X also makes use of a pricing mechanism, or monetary institution, which are social structures. Thus X's "interactive relations" involve both X's relations to other individuals and X's relations to social structures.

What is a social structure? Hodgson does not endorse the idea that social structures lurk in some autonomous metaphysical realm. Instead, he takes them to be "emergent" from interacting individuals:

Some social scientists may uphold that social structures involve more than relations between individuals (where such relations also include social positions)... But this would be mistake. A danger here is to reify social structure as something more than an interacting pattern of individuals, which would exist even if the individuals all disappeared. Social structures are essentially groups of interacting social individuals, possibly including social positions, and with emergent properties resulting from this interaction.<sup>5</sup>

If this is correct, then X's relations to social structures are really multi-place relations holding of the whole population, of which X is a member. These multi-place relations emerge from (and hence supervene on) patterns of one-to-one interaction in the population as a whole. In other words, a first pass at interpreting (b) is:

(b') The social world globally supervenes on individuals and one-to-one interactive relations among the whole population.

In short, Hodgson's thought is that ontological individualism is false if we just consider isolated individuals, but it is true if we add the worldwide distribution of one-to-one interactive relations between people as well.

Hodgson's perspective seems plausible, and versions of it are widely held. However, it has several problems. First, (b') is somewhat redundant. Global supervenience is already a claim about the spread of properties over the whole world, including the whole population. So (b') is the same as:

(b'') The social world globally supervenes on individuals and one-to-one interactive relations.

Second, supervenience is most clearly understood as a relation between properties, not sets of objects. But (a) is written in terms of objects, and (b)–(b'') involve a mix of objects and relations. To make the claims grammatical, we should rewrite them in terms of sets of properties.<sup>6</sup>

---

<sup>5</sup> Hodgson 2007, p. 221. It is not clear from Hodgson's discussion what work is done by "possibly including social positions." See Zahle 2006, pp. 331–334 for an interesting discussion of the debate over "role predicates" in ontological individualism.

<sup>6</sup> Or at least cash out a different notion of supervenience that applies to objects. But converting the relata to properties is easier. A simple trick can be used to include facts, objects, etc., while treating supervenience as a relation between properties. Instead of fact *F*, we can substitute the property *being such that F*. Instead of object *x*, we can substitute one or more of the following: (a) the property *being such that x exists*, (b) the set of intrinsic properties of *x*, (c) the set of properties

Third and by far most significant, what count as the “one-to-one interactive relations”? Claim (a) is not hard to make sense of. To do so, we only need to distinguish intrinsic from extrinsic properties. An intrinsic property I have is *being 6 ft 2 in. tall*; among the extrinsic properties I have are *being taller than Barack Obama*, *being a citizen of the United States*, and *being such that Spain is on the verge of defaulting on its sovereign debt*.<sup>7</sup> A person’s intrinsic properties are those she has in isolation from the rest of the world, properties like her neural states, bodily structure, physical behaviors, etc. (Of course, a person’s neural states might have been *caused by* external factors, but do not ontologically depend on those.)<sup>8</sup> When we talk about “supervening on individuals” in thesis (a), we presumably mean *on intrinsic properties of individuals*. Hodgson asserts, in objecting to (a), that intrinsic properties of individuals do not suffice as the supervenience base for social properties.

However, merely pointing out that social properties depend on some “interactive relations,” as Hodgson does, fails to establish that (a) is mistaken. After all, there are a great many relations between pairs of individuals that themselves supervene on the intrinsic properties on the members of the pair. Consider, for instance, the following three facts:

- (F1) Bob is 4 inches taller than Carol.
- (F2) Bob is 6 feet tall.
- (F3) Carol is 5 foot 8.

Fact (F1) is that the two-place relation *is four inches taller than* holds between Bob and Carol. Facts (F2) and (F3) are that intrinsic properties hold of Bob and Carol, respectively. But of course, there is nothing more to (F1) than the respective facts about Bob and Carol’s heights. That is to say, just because we have added relations on top of intrinsic properties does not mean we have augmented the supervenience base.

Indeed, it is not obvious that even facts like the following fail to supervene on the intrinsic properties of individuals (Hodgson 2007, p. 218):

- (F5) Bob and Carol played a repeated ultimatum game against one another.
- (F6) Bob acquired his cognitive apparatus through socialization and education.

Hodgson, for instance, points out that a game involves norms and rules, and also that the preferences at the outset are socially conditioned. Thus he concludes that we need social relations, not just intrinsic properties of individuals. But this inference is mistaken. Facts (F5) and (F6) might be like (F1), supervening on the intrinsic properties of Bob and Carol, or else on the intrinsic properties of the population as a whole.

---

*being such that x is P*, for all intrinsic properties *P* that *x* has. Thinking in this way has the advantage of forcing us to clarify what we mean when we put “individual people” into the base.

<sup>7</sup> At least, as of mid-June 2012.

<sup>8</sup> On intrinsic vs. extrinsic properties, see Weatherston 2006 and the referenced literature. It also has a nice discussion of relational properties versus extrinsic properties.

That is not to say that (a) is the best interpretation of ontological individualism. In fact, if ontological individualism is to have a chance of being true, the ontological individualist surely needs to include certain extrinsic properties in the “individualistic base.” But an argument like Hodgson’s does not show it. It neglects the crucial task of delineating exactly *which* extrinsic properties and relations must be added to the base, and fails to explain the shortcomings of a base that includes only the intrinsic properties of individuals.

Here are some better reasons for expanding the individualistic base, to include certain extrinsic properties of individuals. Economic exchanges would not occur without bundles of goods, so perhaps *having such-and-such a bundle of goods* should be counted as an individualistic property. Communication between two people cannot occur without a physical medium, so perhaps properties of the air between us should be somehow included among our individualistic ones. Many people think that *having such-and-such a belief* is an extrinsic property, so perhaps extrinsic attitudes should count as “individualistic.”<sup>9</sup> Or perhaps the way to go is to include among the “individualistic” properties not just those that are intrinsic to an individual, but properties that are intrinsic to a pair of individuals acting and interacting with one another.

Any or all of these may be part of an account of the individualistic base. Resolving the “individualistic base” is the key—and perhaps most overlooked—part of clarifying ontological individualism. On the one hand, I have pointed out that arguments like Hodgson’s fail to challenge a very narrow interpretation of the individualistic. They do not say enough about “interactive relations” to show that intrinsic properties of the individuals in the population are inadequate for grounding the social. On the other hand, the examples of exchange, communication, and so on, do suggest that intrinsic properties of individuals are inadequate for grounding the social.

The uncharitable interpreter will take exchange, communication, and so on, to be counterexamples to ontological individualism. The charitable interpreter will look for ways to expand the set of properties that count as “individualistic,” so as to include certain extrinsic properties. Still, even the charitable interpreter cannot just take *every* extrinsic property of individuals to be individualistic. You, for instance, have the extrinsic property *being such that Spain is near default on its sovereign debt*. That is not plausibly an individualistic property of yours. Nor can the ontological individualist assume, as Hodgson does, that “social structures” are emergent properties of patterns of individual interaction. That is little more than a restatement of a thesis of ontological individualism itself. Instead, the ontological individualist needs to clarify which properties count as the individualistic ones, preferably for independent reasons, and then proceed to show that the social properties globally supervene on them.

---

<sup>9</sup>The idea comes from Putnam 1975. Much of Pettit 1993 aims to accommodate an “externalist” theory of attitudes.



### 2.2.3 *The Social*

The third part to clarifying ontological individualism is to pin down what counts as the social world, that is, what is claimed to be “exhaustively grounded” by the individualistic base. Relative to the other two, this is the easy one.

Now, it is surely impossible to give a comprehensive set of the social properties or facts, especially since they change over time.<sup>10</sup> Strictly speaking, the ontological individualist holds that *all* social facts, objects, properties, or events are exhaustively built out of individualistic ones. But if there is any place in the claim where we can tolerate vagueness, it is this one. To draw an analogy, we do not have a clear sense of the set of mental properties, but if we can conclude that beliefs, desires, pains, and so on supervene on neurophysiology, we can develop a reasonable degree of confidence in the more general supervenience thesis. Likewise, even without a comprehensive set of the “social,” if we can conclude that markets, prices, GDP, kinship relations, crime levels, elections, languages, and so on supervene on the individualistic,<sup>11</sup> we can be confident in ontological individualism rather generally.

Probably the best way to approach this part of the thesis of ontological individualism is to look through the social sciences for a nice broad range of social properties and objects. It is also possible to apply ontological individualism to some specific field or domain of the social sciences. An ontological individualist may be happy, for instance, to restrict her claim to the set of entities referred to in the current macroeconomics literature.

And, of course, if someone aims to deny ontological individualism, she does not need to characterize the set of the “social” at all. She only needs one counterexample, a social fact, object, property, or event that fails to be exhaustively grounded by the individualistic. The more the better, but in principle one counterexample would do.

### 2.2.4 *Skepticism About Levels*

As I mentioned at the outset, ontological individualism is a thesis in inter-level metaphysics. It is a claim about the relation between entities at the social level and entities at the individualistic level. In the above discussion, I have spoken of “high-level” and “low-level” entities, and drawn analogies with the mental and the neural, and with the chemical and the microphysical. There are reasons for being skeptical about this hierarchical picture of distinct levels. Jaegwon Kim, for instance, has

---

<sup>10</sup> For instance, new social groups are being formed all the time, as are new types of social groups, new financial instruments, and so on.

<sup>11</sup> Here too, the trick I described in footnote 6 can be used to convert these into corresponding properties.

argued that we should think of levels as increasingly inclusive sets, where the higher levels include everything at the lower levels and more (Kim 2002). William Wimsatt has argued that the sciences are too interlinked to be arranged in levels at all, and that the closest we can get is different scales of aggregation (Wimsatt 1994). Other worries apply specifically to the individual and the social. Philip Pettit has argued that individual attitudes are partly constituted by social entities (Pettit 1993), and many people have argued that individual people as well are “socially constituted” (Foucault 1970 and many others), though often these claims are vague as to whether this is an ontological or a historical claim.

All of these are threats to ontological individualism. If we cannot distinguish the social from the individualistic in the first place, then we cannot clarify the thesis of ontological individualism. Providing satisfactory treatments of the “individualistic base” and the “social facts,” as I discussed in Sects. 2.2.2 and 2.2.3, amount to distinguishing these levels from one another. (Here too, it is more crucial to delineate the individualistic so that it is distinct from the social, than to give a complete account of the social.) If the skeptics about levels are correct, these cannot be achieved, and ontological individualism is no thesis at all.

### 2.3 Anchor Individualism

Nearly all discussion of the role of individuals in “building” the social world has understood it in the terms I have just discussed. There is, however, an entirely distinct thesis about this role: the thesis that social facts are *exhaustively anchored* by individualistic facts.

To introduce the idea of anchoring, consider the recycling bin in front of my apartment. In my town, the bins designated for recycling are large blue plastic cans. Black cans are for non-recyclables. This is not a natural fact about the cans, or a matter of their intrinsic functionality. A black can would work as well for recycling as a blue one. But as it is, when I throw recyclables into the black can or non-recyclables into the blue one, my neighbors scowl at me. And if the city catches me, I am fined.

What makes it the case that blue cans are recycling bins, and black ones bins for non-recyclables? The most widely discussed approach to such questions is John Searle’s treatment of “institutional facts” in *The Construction of Social Reality* (1995) and *Making the Social World* (2010). The reason *being a large blue plastic can* is the condition for *being a recycling bin*, Searle argues, is that we *collectively accept a constitutive rule* for recycling bins. A constitutive rule is a formula like<sup>12</sup>:

(R) If x is a large blue plastic can on the sidewalk, then x is a recycling bin.

---

<sup>12</sup> This is a simplified version. Searle proposes a couple of versions in Searle 1995 and refinements in Searle 2010. See Hindriks 2008, Thomasson 2003, and Epstein 2014, for criticisms of Searle’s formulation.

When the members of a community take a certain attitude toward (R)—in particular, when each member accepts (R) as holding for the community—all the blue plastic cans on their sidewalks count as recycling bins. Notice that, in this account, there are two very different kinds of facts that make the can in front of my apartment a recycling bin. One fact is that the can in front of my apartment is a large blue plastic can. This is the antecedent of (R)'s if. . .then. . . statement. The second fact is the fact about collective acceptance, i.e., that the community collectively accepts (R).

Searle's is one of many theories of how principles like (R) are anchored—i.e., how the conditions for being a social entity like a recycling bin are put in place. In Searle's theory, they are anchored by collective acceptance. Historically, there have been many alternative theories aiming at a similar target. Hume, for instance, proposed that principles like (R) are *conventions*, and gave a rudimentary account of them. A convention is, for Hume,

a sense of common interest; which sense each man feels in his own breast, which he remarks in his fellows, and which carries him, in concurrence with others into a general plan or system of actions, which tends to public utility.

(Hume 1777/1975, p. 257)

Hume is not as explicit as Searle is about distinguishing the fact that a principle like (R) is in place, from the facts that anchor (R). Hume uses the term 'convention' ambiguously, to denote both the beliefs and regularities that give rise to a principle like (R), and to denote the principle itself, that the beliefs and regularities give rise to.<sup>13</sup> But the key place where Searle differs from Hume is above all in his theory of anchoring, and to a small extent in how to think of and formulate principles like (R).

Another alternative is H.L.A. Hart's theory of laws, in *The Concept of Law* (Hart 1962). The (R)-like principles Hart theorizes about are of course laws, not constitutive rules or conventions. His "practice" theory is similar in some ways to Hume's theory, and similar in some ways to Searle's. Like Searle, he insists that we have attitudes toward (R) itself, not just toward regularities of activity. But like Hume, he also takes regularities of practice to be conditions for a principle like (R) to be anchored.

All of these views take social facts and properties to be the products of humans, but in a different way from how the Senate or a soccer team are "composed" of people or "constituted" by people. My recycling bin is not composed of people; it is composed of blue plastic. Nor is it a fiction, or composed of my thoughts or attitudes. My attitudes are weightless, but the recycling bin is heavy. In any of these theories, the two different kinds of facts—i.e., the blue-plastic-can facts vs. the collective-attitude facts, or the blue-plastic-can facts vs. the belief-and-regularity facts—play very different roles.

I suggest we think about anchors in the following way. The facts that anchor principles like (R)—whether collective attitudes, beliefs and regularities, practices,

---

<sup>13</sup> Hume regards his formula as a kind of analysis of what a convention is. Lewis and other theorists of convention also tend to use the term ambiguously, for the same reason.

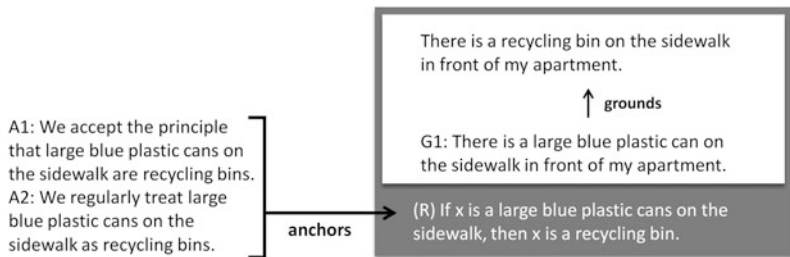


Fig. 2.1 Depicting anchors and grounds for a social fact

or ideas—set up a kind of “frame” or “context.” In that “frame,” social facts are grounded in conformance to principle (R). That is, if (R) is anchored for us, then all it takes for something to be a recycling bin is for it to be a large blue plastic can on the sidewalk. In short, the “anchoring facts” set up principles like (R), and principles like (R) give the grounding conditions, for social facts to be the case.

For example, suppose that Hume’s theory of anchoring is correct: certain beliefs and regularities of practice set up a frame in which, wherever there is a large blue plastic can on the sidewalk, there is a recycling bin. That is, these beliefs and regularities of practice anchor a principle like (R), a principle that is true throughout the frame. Searle calls his particular version of principles like (R) “constitutive rules.” But to broaden the idea to include views such as Hume’s and Hart’s, I will more generically call these “frame principles.” Figure 2.1 depicts the relationships among the various facts I have described, using a Hart-style theory of anchoring as an illustration.

At the left are the facts {A1, A2} that anchor the frame principle (R). That is, according to a Hart-style theory, facts about our attitudes towards (R) and practices conforming to it by members of the community. At the right, the bottom of the box gives the frame principle (R). And within the frame is a set of facts {G1} about a blue plastic can on my sidewalk, which in the frame ground the fact that there is a recycling bin on my sidewalk.

For one more example, consider the fact *Obama is President of the U.S.* In order to be President, one must be at least 35 years old, elected by the Electoral College, take the oath of office, and so on. The frame principle for *being President of the U.S.* is therefore something like:

(P) If x is a person who is at least 35 years old, was elected by the Electoral College in the last 4 years, and has taken the oath of office, then x is President of the U.S.

Here also, we need to distinguish two kinds of facts. There are facts that anchor the frame principle (P), such as historical facts about the U.S. Constitution or our attitudes towards it. And there are facts within our frame that ground the fact that Obama is President of the U.S., such as the fact that Obama is at least 35 years old.

### 2.3.1 *The General Thesis*

In describing ontological individualism, I pointed out that it is a thesis about social facts *in general* being built out of facts about individual people. Similarly, anchor individualism is a thesis about how principles like (R) and (P) are anchored *in general*: namely, by facts about individual people. Searle's collective acceptance theory, for instance, is a theory about how all constitutive rules must be anchored. Whenever a constitutive rule is anchored in a community, it is anchored by the "we-attitudes" of individuals in the community. Searle's basic theory,<sup>14</sup> in fact, gives us a reductive account of the anchors for any given constitutive rule: *Constitutive rule x holds in a community c if and only if every member of c has the we-attitude: We accept x.*

Similar to ontological individualism, however, this is more than we need, if we only want to argue that anchors in general must be individualistic. A theory of convention, for instance, might take there to be a diverse set of facts that together anchor a variety of conventions. It might not regard conventions as being anchored one by one. It may be, for instance, that we have a huge number of conventions about driving cars, and that all of these conventions are jointly anchored by a huge set of interconnected behaviors and beliefs about driving. Just as ontological individualism is the generic thesis that social facts are exhaustively built out of facts about individuals, anchor individualism is the generic thesis that the grounding conditions for social facts are exhaustively anchored by facts about individuals. That is, it is a claim about a relation holding between a set of "social-level principles" and a set of "individual-level stuff." And as with the thesis of ontological individualism, if we want to be more precise about the thesis, we need to clarify three topics.

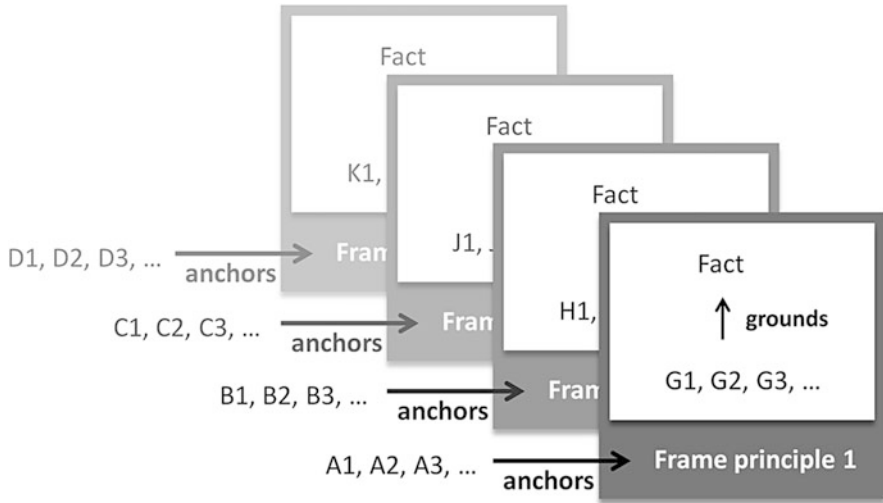
1. What the exhaustive anchoring relation is, that holds between the social-level principles and the individual-level stuff
2. What the individualistic base is: individual people, their mental states, their actions, habits, or practices, various relations among them, etc.
3. What the frame principles are, that are anchored by the base. Are they constitutive rules for institutional facts, social conventions for social kinds, formally enacted laws, or something else?

### 2.3.2 *The Exhaustive Anchoring Relation Between Sets*

The intuitive idea of anchor individualism is simple. Take a set *S* of frame principles for social facts, which might include principles like (R) and (P). We want to say that anchor individualism is true just in case all the principles in *S* are exhaustively anchored by individualistic facts.

---

<sup>14</sup> Searle has modified this a bit in Searle 2010, with his discussion of collective recognition.



**Fig. 2.2** Depicting multiple frames

To specify this more carefully, we can introduce a relation similar to supervenience, but that ranges across different “frames.” The various supervenience relations are used to capture the notion that once the “supervenience base” is fixed, then that fixes the “supervening” properties or facts. Similarly, here we want to capture the notion that if we fix the set of individualistic anchoring facts, then we fix the frame principles in *S*.

Thus we need to recognize that there are many different ways of “framing” the world. In Fig. 2.1, a single frame principle was depicted, along with its anchors. But to make a claim about anchors in general, we need to think about the range of all frames. This is depicted in Fig. 2.2.

Each of the sets of anchors, {A1,A2,...}, {B1,B2,...}, etc., is a set of facts. We might think of {A1,A2,...} as a set of facts holding at one time, such as the year 1950, {B1,B2,...} as holding at a different time, such as 1975, and so on. Or we might think of {A1,A2,...} as holding in one community such as the U.S., and {B1, B2,...} as holding in a different community, such as Mexico. Or we might think of them as holding in different “worlds,” in a possible worlds framework. Or we might think of them as iterated: {A1,A2,...} is a set of facts in the frame anchored by {B1, B2,...}, and {B1,B2,...} is a set of facts in the frame anchored by {C1,C2,...}, and so on.

Each of these sets of facts serves to anchor the frame principles in its corresponding frame. Suppose, for instance, that the following are anchors for two different frames:

A1: We accept the principle that large blue plastic cans on the sidewalk are recycling bins.

A2: We regularly treat large blue plastic cans on the sidewalk as recycling bins.

B1: We accept the principle that small green metal cans on the sidewalk are recycling bins.

B2: We regularly treat small green metal cans on the sidewalk as recycling bins.

Thus {A1, A2} anchors principle (R), whereas {B1, B2} anchors a different principle in that set's respective frame.

Searle claims that the only way to anchor a constitutive rule is by collectively accepting that rule. This means that across all possible frames, facts about collective acceptance are necessary and sufficient for fixing the frame principles. Another way of putting it is this: once we fix the facts about what is collectively accepted in the anchoring context, we have thereby fixed the facts about what constitutive rules are anchored by that context.

This is a way we might begin understanding an "exhaustive anchoring" claim. Similar to a supervenience claim, it asserts that once we have fixed a certain kind of anchoring fact, that suffices to fix a kind of fact anchored by it. Technical machinery for this can be developed, as it has been for supervenience. Clarifying the exhaustive anchoring relation is the first step in interpreting the claim that constitutive rules are *exhaustively anchored* by collective acceptance, or that conventions are *exhaustively anchored* by beliefs and regularity of activity, or more generally that frame principles for social facts are *exhaustively anchored* by individualistic facts.

### 2.3.3 *The Individualistic Base*

Although there are many theories of anchoring in circulation, it is rare to consider the set of facts in general that serve to anchor frame principles. As I described in connection with ontological individualism (in Sect. 2.2.2), some people think about individualistic properties as mental or psychological properties. Among theorists of anchoring, many people (including Searle) think of the anchors for social facts as mentalistic. They regard social facts as a kind of "projection" by the mind onto the natural world. Other views of anchoring, as I mentioned, involve practices or activities as well as mental states. Hume's theory of convention and Hart's "practice theory" of law may be counted among these. As with ontological individualism, it must be decided, to make anchor individualism precise, which properties or facts count as the individualistic ones.

A similar pattern emerges as we saw with ontological individualism. The facts that count as individualistic surely include at least certain mental facts about individual people. If a wholly mentalistic theory of anchoring, like Searle's, is correct, then the thesis of anchor individualism is straightforwardly true as well. On the other hand, Hume's "regularities" and Hart's "practices" are not intrinsic to individual people. Therefore, if one of these theories of anchoring is correct, we have to make a choice. We either need to expand the individualistic base, so as to include practices among the individualistic facts, or else the theory is a counterexample to anchor individualism.

Here too, it is a delicate matter whether to expand the individualistic base. It seems foolish to reject anchor individualism only because we have chosen an excessively strict notion of “individualistic.” On the other hand, it also seems foolish to be so committed to anchor individualism that we will expand the individualistic base indefinitely, to the point that the thesis becomes empty.

### 2.3.4 *The Frame Principles for Social Facts*

The thesis of anchor individualism holds, generically, that *all frame principles for social facts* are exhaustively anchored by individualistic facts. To be precise, it is not the social facts like *The blue plastic can on my sidewalk is a recycling bin* or *Obama is President* that are anchored in a community’s frame. Instead, it is frame principles like (R) or (P). Nonetheless, corresponding to any social fact (like *Obama is President*) is one or more frame principles (like (P)). So an intuitive approach to think about the set of frame principles is to work from the set of social facts discussed in Sect. 2.2.3. Assume that frame principles have a certain form, such as *If facts of type  $X_1, \dots, X_n$  are the case, then fact  $Y$  is the case*. The frame principles for social facts are those that have the social facts from Sect. 2.2.3 in the “Y” position.

As with ontological individualism, anchor individualism can be applied to some specific sub-domain of the social sciences. Indeed, particular theories of anchoring are often held to be theories of some particular sort of frame principle. Searle’s collective acceptance theory is a theory of anchoring for constitutive rules, having a certain form. These constitutive rules are the frame principles for certain kinds of social facts, ones that Searle calls “institutional facts.” Hume’s theory is, of course, a theory of convention. Hume applies his theory to standard examples of social properties, such as ownership, money, and language, but does not say whether all conventions give rise to social facts or institutions, or if only some do, and conversely whether all social facts or institutions are conventional. And Hart’s is a theory of laws.

## 2.4 Comparing the Two

Ontological individualism is a thesis about the grounds of social facts within our frame, and anchor individualism is a thesis about how frames in general (including ours) are anchored. Having sketched these two theses, the natural question to ask is whether they are actually distinct from one another. In particular, are anchors just more grounds? For example, suppose Searle’s theory is correct, and constitutive rule (P) is anchored by collective acceptance of (P) in the community. Then shouldn’t we just say that the fact *Obama is President of the U.S.* is grounded by *both* the facts like *Obama was elected by the Electoral College* and by facts like *We*





**Fig. 2.3** Treating anchors as more grounds

*collectively accept (P)*? This idea is depicted in Fig. 2.3, as an alternative to depictions in Figs. 2.1 and 2.2.

After all, it seems that the collective acceptance fact is just part of the supervenience base of the fact *Obama is President of the U.S.* If we did not collectively accept (P), that fact would not be the case.

This, I propose, is the wrong way to think about the relation between anchors and grounds. The kind of counterfactual I just mentioned—if *we did not collectively accept (P)*, *Obama would not be President*—is more subtle than it seems. In one sense, it can be evaluated as comparing the actual frame with a counterfactual frame. Comparing one frame with another, it is correct to say that if we collectively accepted a different rule (P'), then Obama would not actually be President. However, in a different sense, the counterfactual can be evaluated *inside* our frame. Within a frame, for someone to be President only depends on being over 35, voted in by the Electoral College, and taking the appropriate oath. After all, the whole idea of a constitutive rule, law, convention, or other frame principle is that it gives all the conditions for something to have a given social property like *being President*.

When we assess typical supervenience claims for social facts, we do so inside a frame. Consider, for instance, the property *being a senior citizen*. This is a social property whose conditions are anchored by social facts—perhaps by law, by convention, by collective acceptance, or by practices. The grounds for being a senior citizen are that one be at least 65 years old. When we evaluate the conditions under which a person may or may not be a senior citizen, we consider different situations, in which the person has a different age. Is the person younger than 65, or not? If we wanted to, we might also consider situations in which the instantiation conditions for the property changed; for instance, where the body of U.S. law is different. But that is changing the conditions for *being a senior citizen*. As it is anchored in our frame, the social property *being a senior citizen* supervenes on the age properties of individual people, and that alone.

I do not intend to mount a sustained defense of this claim, in this chapter. I do want to point out, however, that the defender of either form of individualism should be an even more insistent advocate for distinguishing anchors from grounds than I am. Consider how Fig. 2.3 looks, as applied to the example of the recycling bin. This is depicted in Fig. 2.4.

In Fig. 2.4, the social fact *There is a recycling bin on the sidewalk in front of my apartment* is depicted as being grounded by facts about our attitudes together with facts about blue plastic. That is, it is depicted as being grounded by—to put it in

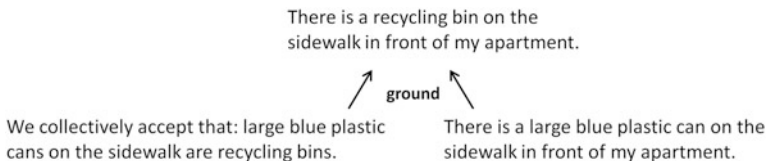


Fig. 2.4 Treating the anchors for (R) as just more grounds

Searle’s terminology—both “intentional facts” and “brute facts.” But Searle would not suggest that the “brute facts” must be individualistic. The key idea of his theory is that individual attitudes project features onto the brute world of non-persons.

The anchor individualist in general has a stake in keeping the grounds like *being a blue plastic can* separate from the individualistic anchors. The typical anchor individualist will have no interest in being an ontological individualist, and will reject the thesis that social facts supervene on individualistic ones. Because whether we collapse the anchors into the grounds, or keep the grounds separate, a property like *being a recycling bin* depends at least in part on the property *being a blue plastic can*.

Conversely, the ontological individualist will take issue with the perspective of anchor individualism. The ontological individualist insists that non-individualistic facts have no place in grounding social ones. This means that she may restrict the set of social facts, so as to exclude facts like *There is a recycling bin in front of my apartment*. She may be an ontological individualist about macroeconomic facts, for instance, but not about recycling facts. Or else she may expand the individualistic base, so that a property like *being a large blue plastic can in front of my house* is an individualistic fact about me or other people.

## 2.5 Conclusion

To many people, both ontological individualism and anchor individualism seem to be appealing for the same reason. Both seem to deflate worries about the “reality” of the social world. The social world is just *us*, both theses hold. However, the two theses deflate these worries in conflicting ways. For an ontological individualist, the prototypical example of a social fact is one about a group, like a court or legislature, which is composed of individual people. The ontological individualist typically regards social facts as emerging from interactions among individual people, in combination with one another. For an anchor individualist, in contrast, the prototypical example is a fact about dollar bills, recycling bins, or boundaries made of lines of stones.

Advocates of methodological individualism will favor ontological individualism. Although ontological individualism does not imply methodological individualism, it is at least a first step. If facts about a mob are exhaustively grounded by

facts about individual people, then there is a chance that we can explain those facts in terms of facts about individual people.

Anchor individualism, on the other hand, is not as congenial to methodological individualism. Suppose, for instance, that one is an anchor individualist but not an ontological individualist about the fact *There is a recycling bin in front of my apartment*. That is, this fact does not supervene on individualistic facts, but the frame principles for recycling bins, apartments, etc., are exhaustively anchored by individualistic facts. When we are explaining this fact—giving a causal or other kind of explanation for why it is the case—its supervenience failure means that it is very unlikely that that explanation can be given entirely in terms of facts about individual people. Instead, the explanation will in all likelihood need to involve some of the non-individualistic facts that actually ground it. Some people seem to think that if constitutive rules (or other frame principles) for money or recycling bins are anchored by attitudes, then money and recycling bins are just “in the head,” like fictional characters in a mental play. But again, recycling bins are heavy, and made of plastic. Even if their frame principles are anchored by attitudes, facts about them are grounded by facts about things other than attitudes.

Altogether, ontological individualism and anchor individualism are very different from one another. Discussions in the metaphysics of the social world are often enormously confused, because of the failure to notice this difference, or to see its magnitude.

## References

- Bennett, K. (2011). Construction area: No hard hat required. *Philosophical Studies*, 154, 79–104.
- Correia, F. (2005). *Existential dependence and cognate notions*. Munich: Philosophia Verlag.
- Currie, G. (1984). Individualism and global supervenience. *British Journal for the Philosophy of Science*, 35, 345–358.
- Epstein, B. (2008). When local models fail. *Philosophy of the Social Sciences*, 38, 3–24.
- Epstein, B. (2009). Ontological individualism reconsidered. *Synthese*, 166(1), 187–213.
- Epstein, B. (2011). Agent-based models and the fallacies of individualism. In P. Humphreys & C. Imbert (Eds.), *Models, simulations, and representations* (pp. 115–144). New York: Routledge.
- Epstein, B. (2014). Social objects without intentions. In A. Konzelmann Ziv & H. B. Schmid (Eds.), *Institutions, emotions, and group agents: Contributions to social ontology*. Dordrecht: Springer.
- Fine, K. (2001). The question of realism. *Philosopher's Imprint*, 1(1), 1–30.
- Foucault, M. (1970). *The order of things: An archaeology of the human sciences*. London: Tavistock Publications.
- Hart, H. L. A. (1962). *The concept of law*. Oxford: Clarendon.
- Hindriks, F. (2008). The status account of corporate agents. In H. B. Schmid, K. Schulte-Ostermann, & N. Psarros (Eds.), *Concepts of sharedness: Essays on collective intentionality* (pp. 119–144). Frankfurt: Ontos.
- Hodgson, G. (2007). Meanings of methodological individualism. *Journal of Economic Methodology*, 14(2), 211–226.

- Hume, D. (1777/1975). *Enquiries concerning human understanding and concerning the principles of morals* (L. A. Selby-Bigge, Ed.). Oxford: Clarendon Press
- Kim, J. (1998). *Mind in a physical world*. Cambridge: MIT Press.
- Kim, J. (2002). The layered model: Metaphysical considerations. *Philosophical Explorations*, 5 (1), 2–20.
- Kincaid, H. (1986). Reduction, explanation and individualism. In M. Martin & L. C. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 497–515). Cambridge: MIT Press.
- Lukes, S. (1968). Methodological individualism reconsidered. *British Journal of Sociology*, 19, 119–129.
- McLaughlin, B., & Bennett, K. (2005). Supervenience. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Accessed 1 November 2011. URL = <http://plato.stanford.edu/archives/sum2010/entries/supervenience/>.
- Pettit, P. (1993). *The common mind*. New York: Oxford University Press.
- Putnam, H. (1975). *The meaning of 'meaning'* (Philosophical papers). Cambridge: Cambridge University Press.
- Rosen, G. (2010). Metaphysical dependence: Grounding and reduction. In B. Hale & A. Hoffman (Eds.), *Modality: Metaphysics, logic, and epistemology*. Oxford: Oxford University Press.
- Schaffer, J. (2009). On what grounds what. In D. Manley, D. J. Chalmers, & R. Wasserman (Eds.), *Metametaphysics: New essays on the foundations of ontology*. Oxford: Oxford University Press.
- Searle, J. R. (1995). *The construction of social reality*. New York: Free Press.
- Searle, J. R. (2010). *Making the social world: The structure of human civilization*. Oxford: Oxford University Press.
- Thomasson, A. (2003). Foundations for a social ontology. *ProtoSociology*, 18–19, 269–290. Accessed 1 November 2011. URL = <http://plato.stanford.edu/archives/fall2008/entries/intrinsic-extrinsic/>
- Weatherston, B. (2006). Intrinsic vs. extrinsic properties. In *Stanford encyclopedia of philosophy*.
- Wimsatt, W. (1994). The ontology of complex systems: Levels of organization, perspectives, and causal thickets. *Canadian Journal of Philosophy* 20 (Supplement), 207–274.
- Zahle, J. (2006). Holism and supervenience. In S. P. Turner & M. Risjord (Eds.), *Philosophy of anthropology and sociology* (pp. 311–342). Amsterdam: North Holland.

## Chapter 3

# Social Entities and the Basis of Their Powers

Dave Elder-Vass

**Abstract** This paper offers an emergentist justification for the claim that social structure is causally significant when it takes the form of social entities with *relationally* emergent causal powers. Such powers are generated by processes of interaction between the characteristic set of parts, given the characteristic relations between them, that occur in entities of the type concerned. The paper offers a justification of this argument in the face of criticisms that it is too weak to ground causal claims, arguing on the contrary that debates in the philosophy of mind have raised expectations of emergence theory beyond what it can plausibly deliver. The relational form of emergence theory provides a viable refutation of eliminative reductionism as a generalised strategy, and can also be employed to refute the form of eliminative reductionism known as methodological individualism. This kind of emergence theory delivers just what we need from the concept: it justifies the need for higher level sciences to study higher level mechanisms and powers, mechanisms whose explanation will never be made redundant by some lower level theory of everything. It also supports a specific way of thinking about social structure that is arguably rather different from the ways that have tended to dominate sociological discourse. My project includes developing such theory, and the final part of the paper illustrates how this way of thinking about emergence in the social sphere leads to useful and interesting ways of reconceptualising social structure.

---

D. Elder-Vass (✉)

Department of Social Sciences, Loughborough University, Loughborough, UK  
e-mail: [d.elder-vass@lboro.ac.uk](mailto:d.elder-vass@lboro.ac.uk)

### 3.1 Introduction

This paper offers an emergentist justification for the claims that certain sorts of social entities are causally significant, and that we should therefore reject methodological individualism, understood as a denial of the possibility of such structural causation.<sup>1,2</sup> It argues that social structure is causally significant when there are social entities with emergent causal powers. Such powers are generated by processes of interaction between the characteristic set of parts of such an entity and the characteristic relations between them that occur in entities of the type concerned. The paper will offer both a justification of this argument and some brief illustrations. This argument, however, should also lead us to reject certain sorts of emergence theory, including much of the thinking about emergence that has been developed within the philosophy of mind. And it should lead us to question those forms of social structural discourse that *assume* structures are causally significant without *explaining* how the structures concerned can be significant in terms that are consistent with a coherent account of emergence.

Because the form of emergence theory advocated here is sometimes denigrated by philosophers on the grounds that it is too weak to ground causal claims, the paper begins by challenging the debate in the philosophy of mind that has conditioned them to expect something more from emergence theory than it can reasonably deliver. It then continues by outlining in general terms the *relational* form of emergence theory that I employ and making clear what can be claimed for it: that it provides a viable refutation of eliminative reductionism as a generalised strategy, where eliminative reductionism is defined as the denial of the causal significance of a category of entities or properties on the grounds that the causal work concerned is really done by some lower level category of entities or properties. In particular, I claim that this form of emergence theory can be employed to refute the form of eliminative reductionism that has come to be known as methodological individualism in the social sciences. The paper argues that this is a kind of emergence worth having, and indeed a kind of emergence that delivers just what we need from the concept: it justifies the need for higher level sciences to study higher level mechanisms and powers, mechanisms whose explanation will never be made redundant by some lower level theory of everything.

This version of emergence theory supports a specific way of thinking about social structure that is arguably rather different from the ways that have tended to dominate sociological discourse. My project includes developing such theory; it is ultimately a sociological and not a philosophical project, and the final part of the paper illustrates how this way of thinking about emergence in the social sphere

---

<sup>1</sup>I would like to thank Julie Zahle for her invitation to present this paper at the workshop ‘Individualism, Holism, Explanation, and Emergence’ in Copenhagen, November 2012, and the participants in that workshop for their stimulating questions and contributions.

<sup>2</sup>The paper draws heavily on arguments from my recent work (particularly Elder-Vass 2010).

leads to what I take to be useful and interesting ways of reconceptualising social structure.

## 3.2 Emergence and the Philosophy of Mind

At the heart of my approach to emergence is the idea that there are things in the world (objects, or entities, if you like) that exert a causal influence and that emergence theory is about the relation between the causal significance of things and that of their material parts. At its simplest, this kind of emergence theory claims just that things have causal powers that their parts would not have if they were not organised into this kind of whole. Such emergent properties are thus similar to the concepts of *collective* or *systemic* properties.

Philosophers sometimes dismiss these as familiar and uninteresting, as if they were entirely unimportant by comparison with stronger forms of emergence. The pursuit of such forms is heavily focused on the philosophy of mind, and the search for a theory of emergence that will justify the claim that *mental properties* are special and in particular that they are in some sense causally autonomous of *physical properties*. I have gradually come to the conclusion that the entire debate on emergence in the philosophy of mind is shaped by a fundamental error that we may call *residual Cartesianism*: the belief that mental properties are radically different from other kinds of properties and thus in some sense exempt from the laws of causality.<sup>3</sup> Although the physicalist position within the debate is explicitly opposed to this belief, there is a sense in which it seems to be trapped within a terminology and a set of base assumptions that have already been distorted by an interlinked series of Cartesian moves.

The first of these moves is to abandon the notion of things or objects (which I shall call *entities*) and instead to develop the argument in terms of various kinds of *properties*. This is more or less a forced move if one is to defend residual Cartesianism, since it is now generally agreed that there is no such *thing* as a mind (whereas there clearly is such a thing as the brain), and so the mental and the concept of mind are rescued by using these terms to refer to a set of *mental properties*. Connected to this move, the relations between entities are also relabelled as just another variety of properties.

The second move is to separate mental properties from physical properties, a move that Searle has labelled *property dualism* (Searle 2002), and again a necessary move if one is to preserve the idea that mental properties could be fundamentally different from others. This notion of a physical property is however problematic. There is something rather odd, for example, about the fudge of saying that physical properties are properties studied by physics and then immediately going on to treat

---

<sup>3</sup>The term *residual Cartesianism* has been used elsewhere in the literature, although generally with a different residue in mind (e.g. Allen et al. 2012).

neurological properties – which generally are not studied by physics – as physical properties. Taking a different approach, we can usefully talk about physical *entities*, in the sense of material entities, but it would be equally odd to talk about *properties* being physical in the sense of material. What is material about being green, for example? In practice, then, the concept of a *physical property* in this debate simply means “all properties that are not mental properties”: it is a fake other that serves only to legitimate the idea that mental properties are *different* without providing any substantive specification of what it is that they are taken to be different from. Indeed, there seems to be nothing that so-called physical properties have in common that they do not also share with mental properties.

But it would be laughable today to insist that mental properties are entirely unconnected to physical entities in the form of brains, and so residual Cartesianism must make a third move that somehow recognises this relation while nevertheless asserting the independence of mental properties. This is achieved by treating the physical as a *lower level* with an ambiguous relationship to *higher level* mental properties: one in which the nature of these levels is left opaque, and some mixture of dependence and autonomy between them is asserted. This set of moves is not only assumed by emergentists, but also by many of those who question the emergence of the mental. It is built into both the programme and the language of a significant proportion of the philosophy of mind – or at least of that perhaps rather limited section of the philosophy of mind with which I have engaged.

I suggest that all three of these moves are errors. The properties that are of interest in debates on emergence are always properties *of* specific entities or types of entity, and it matters to these debates *which* entities they are properties of. These features are obscured by ignoring entities. Similarly, the relations between entities are quite a different class of phenomenon than properties that are possessed by a single entity in isolation from others. Again, this is a distinction that matters to this debate, as we will see, and it is one that is obscured by the exclusive focus on properties. Without distinguishing between entities, their properties, and the relations between them, I will suggest, we cannot even make sense of the problem of reductionism, let alone possible solutions to it.

It is not unreasonable to suggest that there are phenomena it is useful to label as mental properties: beliefs and desires, for example. But, like Searle, I am baffled by the idea that these are somehow distinct from so-called physical properties (Searle 1992: chapter 1, 2002). For me, mental properties are a subset of the properties of material entities in much the same way that colour properties or electro-magnetic properties are a subset of the properties of material entities. Once we accept this, there is no need to abandon the language of entities and their properties, and no need to struggle with the underpinnings of the separation between mental and physical properties. A mental property is a property of a material entity, a person, just as the capacity to walk or sing or a certain skin colour may be a property of a person. None of these properties is any more or less *physical* than any of the others, and there is therefore no reason for us to find the relation between mental and physical properties as such problematic.



Nor is there any reason to think that mental properties are higher level than other ('physical') properties in general. The concept of levels is meaningful within an entity-based ontology, since there is a compositional hierarchy within any given entity, and it is useful to think of wholes as being at a higher compositional level than their immediate parts, those parts being at a higher level than their own parts, and so on. Given such an ontology, we may call the properties of a given whole *higher level* properties than the properties of its parts. But without such an ontology, it is not at all clear what it would *mean* to call a set of properties *higher level*. Given an entity-oriented ontology, we could say that mental properties, being properties of whole human beings, are higher-level properties than the properties of, for example, neurons, but *not* that they are higher-level properties than other ('physical') properties of whole human beings.

There does, nevertheless, remain a question to which some form of emergence theory might be relevant: What is the relation between a person's beliefs, desires, etc., and that person's neurological structure? Unlike the debate over mental and physical properties, as it has usually been conceived in the philosophy of mind, this is much the same sort of question as many others we may encounter in scientific contexts, such as "What is the relation between the radiation that may be generated by a star and its physical structure?" or "What is the relation between the capacities of an organisation and its structure?" In some respects, all of these questions are scientific, as opposed to philosophical, questions. I am one of those optimists who believe that eventually, if we don't destroy ourselves first, humanity will develop satisfying, useful, and largely accurate scientific explanations of all of these questions. What, then, might be the relevance of emergence theory?

It is useful, primarily, as a response to eliminative reductionists: those who question a whole set of explanations on the grounds that they appeal to the causal significance of a category of entities that they believe to be causally epiphenomenal, usually because they believe that the causal work is really done by some lower level category of entities. This is arguably a problem of the immature sciences. Practitioners of the mature sciences rarely need to worry about eliminative reductionism since in these sciences it is clearly useful to proceed as if entities at a variety of compositional levels have causal significance and there is no obvious reason to doubt that this is so. But sociologists need to worry about it because it is still considered plausible by some to argue that social structures are not causally significant and therefore that we should reject any causal explanation that invokes them. And perhaps psychologists need to worry about it in an age that threatens to become neurologically reductionist. It is therefore useful to have a general account of emergence that may be used in response to such reductionisms and applied to specific cases.

### 3.3 Relational Emergence Theory

As an introduction to such an account, consider what we might need by way of a response to eliminative reductionism. Once we discard residual Cartesianism, there is no need to seek to sustain its continued yearning for realms of existence that are in some sense exempt from causal explanation and yet at the same time capable of exerting a causal influence. We can therefore reject another deeply ingrained philosophical presumption, one that is wrapped up in many uses of the concept of reductionism. This is the assumption that if we can offer a ‘lower level’ explanation for something, then that something cannot itself have causal significance.

One argument that is sometimes deployed in this context is Kim’s causal exclusion argument (e.g. Kim 1993: 203–209, 2006: 558). Kim argues that if there is a mental property and a physical property that realises it, and if we accept that the original physical property causes a change to a new physical state, then we cannot also believe that the original mental property causes a change to the corresponding new mental state. In some versions of the argument, he makes clear that the physical property he has in mind is what he calls a *microstructural* property, which corresponds to a “unique complete microstructural description: that is, any physical system can be exhaustively described in terms of (i) the basic particles that constitute it. . . ; (ii) all the intrinsic properties of these particles; and (iii) the relations that configure these particles into a structure” (Kim 1999: 6). A microstructural property, in other words, is a translation into property language of the concept of an entity: the microstructural property of a system is a complete description of all the parts and relations that constitute it as a system, and so in specifying a microstructural property we are essentially describing the entity concerned. My summary of his argument, then, could be rephrased as follows: if there is a mental property that is a property of a material person in a given state, and if we accept that the original state of the material person causes a change to a new state of the material person, then we cannot also believe that the original mental property causes a change to the corresponding new mental state.<sup>4</sup>

Whatever this argument implies, it does *not* imply that higher level properties, in the compositional sense of *levels* advocated above, are not causally effective if lower level properties are. As Kim points out, microstructural properties are not lower level properties but macro properties: “macro since it belongs to the system as a whole constituted by the system’s basic micro-constituents, their intrinsic properties, and the relations that structure them into a system with unity and stability as a substance” (Kim 1999: 7). Only the whole system, the whole entity in my terms, can have the microstructural property, so it is not a lower-level property in the compositional sense of *levels*. His causal exclusion argument is not about the relationship between the causal power of entities at different compositional levels, but about the relationship between the causal power of an entity at a

---

<sup>4</sup>The reader may substitute ‘brain’ for ‘material person’ if this makes the argument more accessible. But mental states may depend on our entire nervous system and not just our brains.

given level and a property at the same level. Indeed Kim himself appears to advocate a view that is similar to the account of emergence offered in the current paper, though he does not *call* it emergence: “the fact that we can micro-structurally explain why a micro-based property has a certain set of causal powers does not mean that these causal powers are identical with the causal powers of its micro-constituents. Micro-reductively explainable causal powers may be new causal powers, net additions to the causal structure of the world” (Kim 1998: 117). Kim’s argument therefore cannot be invoked against relational forms of emergence theory such as that advocated here.

Let us turn, then, to the ontological basis of relational emergence theory. My base assumption is that our universe is populated with stuff that interacts and as a result forms progressively more complex structures, which sometimes have a degree of stability and persistence, and which in turn interact in progressively more complex ways. It is populated, as a result, by structured entities that we can think of as decomposed hierarchically into parts and sub parts at various levels. Those complex structures can have effects that the same stuff cannot and does not have when it is not organised into such structures. In the sense of emergence advocated here, this means that entities at each level can have emergent causal powers: powers to affect the world that would not be possessed by the parts if they were not organised into entities of this nature.

It is the interactions of the part-stuff that produce the effects of these structures but those interactions depend on the set of relations between the parts that only exist when this structure is present, in other words when the set of relations that is required to organise these parts into this kind of structure is operative. For this reason, I refer to this as *relational* emergence theory. In a sense, it is the relations between the parts of any given whole that provide a bridge or intermediate level between one compositional level of entities and the next: this is a further input into the properties of the whole that is not in itself explained when we explain the intrinsic (that is non-relational) properties of the parts.

In principle, science may provide explanations of how these interactions produce the emergent causal powers concerned; that is, it may identify the *mechanisms* that generate the higher level powers. Such explanations, however, are not reductions in what I take to be the core (that is, eliminative) sense of the term since they do not entail that the higher level structure is not required for the production of the effect. At the risk of confusing things by invoking a term that is often thought to have eliminative connotations, we could call these explanations *explanatory reductions*, but they do not entail the eliminative reductionist conclusion that the whole entity is somehow causally irrelevant.

A common response to such arguments by those of an eliminative reductionist frame of mind is that in such cases it is ‘really’ the parts that are doing the causal work. Let me call this position *reallyism*. The incoherence of reallyism is easily exposed by applying it recursively. If, for example, it were correct to argue that it isn’t ‘really’ an organisation that is exerting a causal influence, but the organisation’s members, then the same conceptual frame could be applied to the members themselves and would lead to a series of further conclusions: that it isn’t really the

members but their cells that are doing the causal work, not really the cells but their molecules, not really the molecules but their atoms, and so on until all causal power drained away into the bottomless pit of our scientific understanding of the most fundamental structures underlying the objects of particle physics (if indeed there are such structures) (cf Block 2003). To sustain the kind of claim advanced by many eliminative reductionists – that *some* structured entities have causal powers and others do not – we would need a further ontological analysis that provides a justification for discriminating between forms of structure that may and may not have such powers. It is rare to find any such argument from realists, or indeed any recognition that their position requires it. One exception is the social psychologist Rom Harré, who argues that humans occupy “two realities. . . biology and conversation”, characterised by causal and symbolic relations respectively (Varela and Harre 1996: 317), and that in each reality all explanations can be reductively expressed in terms of one and only one kind of “powerful particular”. In the former space, the powerful particulars are electrons (or fields), and in the latter they are persons (Harré and Bhaskar 2001: 31). While I find Harré’s account unsatisfactory, he does at least strive to be consistent with the internal logic of reductionist arguments. More typically, realists appear to base the belief that certain sorts of entities (generally people) must be causally significant, despite the logical structure of their own argument, on little more than an anthropocentric prejudice.

My general ontological argument, then is that entities have emergent causal powers when they are capable of exerting an influence on the world that their parts would not be able to exert were they not organised into such a whole.<sup>5</sup> Such powers are produced by mechanisms, processes in which the parts of the entity interact to generate the influence concerned, and we may be able to explain such mechanisms scientifically. This, however, does not alter the fact that these powers would not exist if the whole did not exist, and therefore we may conclude that these are causal powers of the whole and not of the parts. Without such a move, it is difficult to see how we could locate causal power anywhere in the world; with it, we have a coherent way of examining both how causal power is generated and how powers interact to produce events.

In terms of the philosophical literature, then, I am advocating what would usually be labelled a rather weak form of emergence. Stephan, for example, distinguishes between *strong* forms of emergence theory, such as Broad’s, which assert that emergent properties cannot be explained in terms of lower level parts and the relations between them, and *weak* forms, which do not make this claim (Stephan 2002: 79).<sup>6</sup> It is strong forms that are sought by residual Cartesians; whereas in Stephan’s terms relational emergence is a variety of the weak form. Stephan, however, goes on to argue that weak forms are “compatible with reductionistic approaches without further ado” (Stephan 2002: 79). Such arguments must be

---

<sup>5</sup> This argument draws significantly on the early work of Roy Bhaskar (Bhaskar 1975; Elder-Vass 2005).

<sup>6</sup> Other writers have given somewhat different meanings to these terms (e.g. Bedau 1997).

carefully qualified by reference to the *kind* of reductionism at issue. I am tempted to suggest that we might also usefully distinguish between *weak* and *strong* forms of *reductionism*: weak reductions are simply explanations of a property of an entity, whereas strong reductions are explanations of such a property made entirely in terms that are compatible with the non-existence of the whole structured entity. Weak reductions do not entail that the property being explained or the entity possessing the property can be eliminated from a viable explanation of the effects of the property; whereas strong reductions do: they are eliminative reductions in the sense defined earlier. Relational emergence theory, we may then say, is thoroughly compatible with weak reductions of emergent properties but not with strong reductions of emergent properties (although it is compatible with strong reductions of non-emergent, i.e. resultant or aggregative, properties) (Wimsatt 2000).

Philosophers of mind are prone to question the value of emergence theories that are compatible with weak or explanatory reductions, since they appear to be in search of an emergence theory that will exempt mental properties from ‘physical’ explanations. But once we discard residual Cartesianism we have no need for strong theories of emergence. What we *do* need is ontological theories that are compatible with the successful practices of actual science, and thus with us living (as we do) in the kind of world in which such practices *can* be successful.<sup>7</sup> In particular, we need theories that are compatible with two key features of that practice. First, a vast range of entities is treated as having causal significance; and second, this causal significance is taken to be compatible with the production of explanations of how it arises (Gell-Mann 1995: 112). The kind of emergence theory advocated here meets both of these requirements, and provides us with an ontological framework that recognises the need for sciences of each level of structure: sciences that recognise which macro structures have which kinds of causal influence and also seek to explain how they can have it. This is all we need from the concept of emergence. And this gives us all that we need by way of resisting eliminative reductionism: it enables us to justify the assertion that higher level entities have causal powers while resisting the anti-scientific insistence that such powers are in some sense uncaused or unexplainable. This kind of emergence theory may be of no use to those seeking to substantiate pseudo-Cartesian dualisms in defence of concepts of mind that are somehow to be exempted from the normal processes of causality and explanation. But however it is labelled, it is *strong enough* to support a coherent approach to causal explanation.

---

<sup>7</sup> I take this to be one of the core arguments of Bhaskar’s *Realist Theory of Science* (1975).

### 3.4 Causal Powers of Social Entities

My interest in emergence theory, ultimately, is not philosophical. The focus of my work lies in the area where the ontology of the social world meets empirically relevant sociological theory. For me, then, emergence theory is a tool for making sense of concepts like social structure and agency, in ways that can be productive for sociological theory. Does it, for example, provide us with a coherent response to approaches like methodological individualism, a response that works not just as a philosophical argument, but because we can use it to develop accounts of social structure that show *how* it can be causally significant? Developing such accounts has been the primary focus of my work so far, and I hope to show that this does indeed offer at least part of a refutation of methodological individualism. But it also has disruptive implications for previous understandings of social structure, which are at least as important.

Perhaps the simplest way to apply emergence theory to these issues would be to assert that social structures, as they are already understood in existing sociological work, are, or are produced by, emergent properties. Such an approach would bolt on an emergentist justification to an existing body of thinking about social structure in a superficially non-disruptive way, but it would also generate a range of challenging questions. The most striking of these arises from the claim that emergent properties, as I develop the argument, are properties of entities that arise from the interactions between their parts. If this is so, then a coherent bolting on of emergence theory would require us to identify what entities, parts and interactions are involved in existing conceptions of social structure. While there are many ways in which one could attempt to fit these concepts together (Elder-Vass 2010: chapter 4) perhaps the mainstream assumption has been that the entities involved in social structure are *societies*. We might then regard social structures as emergent causal powers of these societies. But there are a number of reasons why such an approach is unsatisfactory.

The most obvious is that the concept of a society is itself rather contentious. It carries with it a series of assumptions, beginning with the assumption that territorial states are taken to be boundary-defining for societies – an assumption that may be labelled *methodological nationalism* (Chernilo 2007). But if this is the case, and if it is societies that are the causal agents at work when social structure is invoked as an explanation, this would seem to imply that social forces should operate similarly throughout the territory of any given state, and differently in the territories of other states. This is problematic because the geographical scope of specific social structural forces clearly varies enormously. Many are influential in spaces that are much larger than the territory of particular states, such as the normative standards espoused by various world religions, or the influence of Microsoft on practices of interaction between humans and computers. Many others are influential in much smaller spaces, such as separatist political parties and very local cultural traditions such as some language dialects. And still more are influential in spaces that are simultaneously both larger in some dimensions and smaller in others than the

territory of particular states, such as the New Age or hip-hop subcultures. If social structures were an emergent property of whole societies, such variations in scope of influence would be incomprehensible, as would be the widespread phenomenon of conflicting social structures within the territory of the same state – employers organisations and trade unions, or meat-eating and vegetarianism, for example.

It is far more plausible to argue that the entities at work when we talk of the causal influence of social structure are not societies at all, but smaller (in most cases) and more specialised social entities. Rather than a few monolithic ‘societies’ exercising social causative powers, what we find is a vast range of what we may call meso-level social entities (Stones 2005). Perhaps the most obvious type of such a social entity is organisations, and we may usefully consider how the causal model advocated here alters theoretical understandings of organisations (Elder-Vass 2010: chapter 7).<sup>8</sup> An organisation is a social entity, whose parts are primarily people, in a structured set of relationships to each other. Those relationships are often characterised as the roles of the individuals concerned. As a result of its members being committed to interact in the ways specified in their roles, the organisation has the capacity to have a causal impact on the world that its members would not have if they were not parts of the organisation concerned. Thus, for example, an orchestra has the causal power to produce harmonious music, a power that is generated as a result of the musicians (and arguably their instruments) who are its parts interacting in the ways specified in their roles (violin player, pianist, conductor, etc.). If they were not organised into an orchestra, the players and instruments would not have the causal power to produce harmonious music.

However, the effects of the relational model of emergence are rather more radical if we consider a different class of social entities: the group of entities I have called *norm circles*. Norm circles, I argue, are the type of social entity that is causally responsible for normative social institutions or social practices. A social practice may be defined as a recognisable pattern of behaviour or action that occurs repeatedly in a social space, ranging from something as simple as standardised forms of greeting to something as complex as the practices that surround (and indeed create) what we think of as property. Not only simple day to day practices like queuing, but also phenomena like money, religion, language, culture and indeed organisations depend on normatively standardised practices.

Sociologists often invoke the concept of a normative social institution as a kind of explanation of the regularity of such practices. What is generally accepted is that such institutions are driven by normative pressures (socialisation, for example) which encourage people to conform to the practice and may penalise those who do not. What is not generally accepted, and indeed has been the focus of over a century of debate, is just what form social institutions take that gives them the capacity to exercise such an influence.

---

<sup>8</sup>The remainder of this paragraph and the following six are drawn largely from Elder-Vass (2012b).

The hypothesis examined in my work is that social practices are produced primarily by the causal power of social entities that I call norm circles (Elder-Vass 2010: chapter 6). A norm circle is the group of people that are committed to endorsing and enforcing a specific norm, a specific standard of observable behaviour. The relation between them that gives them the collective capability to influence behaviour – a greater influence than an unconnected group of individuals would have – is the sense of shared commitment they have to supporting the norm. The members of a norm circle may be unaware of the full extent of the group, and they may not even think of it as a group, but they are generally aware when they act in support of a norm that they are not simply expressing a purely idiosyncratic personal attachment to a particular standard of behaviour. Rather, they are aware that when they do so they are endorsing a standard that others also endorse, and often do so with the expectation that others would support and approve of their action. The individual, in other words, has a sense, however vague and minimal, that she is acting on behalf of something wider than herself when she acts in support of a norm, and that sense increases the likelihood that she will act in its support, by comparison with the isolated individual with a purely personal attachment to the standard of behaviour concerned.

This sense, in turn, is a product of the same process that tends to encourage conformity to the norm – the generative mechanism that underpins the power of a norm circle to increase such conformity. The heart of this process is repeated exposure of individuals to acts of endorsement and enforcement of the norm concerned. If, for example, I repeatedly see people criticising those who try to jump queues, I will start to understand the norm of queuing, and to believe that I face an environment in which I will be sanctioned negatively if I fail to observe it. I will, in other words, develop beliefs about my normative environment which will tend to lead me to conform to the norm of queuing in the future, as a result of the actions of members of the norm circle for queuing. Here, then, social structure – the norm circles that produce the normative environment – is exerting a top-down influence on individual action.<sup>9</sup> Similar effects can be produced without us forming conscious beliefs; most of us, for example, understand and implement the norms prevalent in our social space regarding how close one should stand to someone when talking to them, even though these norms are rarely stated explicitly and are mostly endorsed and enforced rather subtly by non-verbal signals.

Norm circles, then, operate through individuals. On the one hand, it is the actions of individual members of the norm circle acting in support of a norm that signal the normative environment to other individuals and these individual actions, therefore, reproduce and/or transform these social structures.<sup>10</sup> On the other, those pressures do not lead directly and mechanically to norm conformity but rather influence the stored beliefs and dispositions of the affected individuals, which then in turn

---

<sup>9</sup>This is therefore equivalent to the first half of Margaret Archer's morphogenetic cycle of interaction between structure and agency (Archer 1995).

<sup>10</sup>As in the second half of Archer's morphogenetic cycle.



influence their subsequent behaviour. Nevertheless, I argue, the resulting increase in the tendency of those affected individuals to conform to the norm is causally influenced by the norm circle, and not just the individuals. We may compare this with the case of a person who switches on a torch by pressing the switch with her finger: although her finger presses the switch, it could not do so except as a result of the causal influence of the whole individual, and so the event of switching on the torch is caused by a power of the whole individual and not just of the finger. In a similar way, the norm circle can only influence us through its individual members, but those individual members would not influence us in that way, or at least not as strongly and as often as they do, if they were not part of a wider norm circle, and so their act of influencing us is produced by a causal power of the norm circle and not just of the individual. Just because the causal power of the norm circle is exercised through individuals, this does not mean that it is *really* a causal power of individuals, since it is a causal power that would not exist if those individuals were not organised into a norm circle.

There is much more to be said about the theory of norm circles. For example, norm circles in contemporary societies are diversely intersectional – different norms are supported by different but profusely overlapping groups of people. This in turn makes normative change more likely than in more homogeneous societies, since individuals are open to the influence of competing norm circles and may move between them. One implication is that the theory of norm circles is not merely concerned with the reproduction of a stable normative environment. It seeks to explain how normative influences contribute to the production of social actions that conform with prevailing practices, and thus may contribute to the reproduction of the normative environment, but there are many reasons why norms may be transformed rather than reproduced in some social situations. Another feature of this approach is that it is compatible both with cases in which individuals conform with and/or endorse norms as a result of internalising a strongly moral sense that they are right, and also with cases in which they do so for much more instrumental reasons, such as seeking approval and avoiding punishment.<sup>11</sup> Such arguments may also be extended to a variety of other normatively shaped phenomena, such as discourse, language, and knowledge (Elder-Vass 2011, 2012a).

For the purposes of this paper, however, the relevance of norm circle theory is that it illustrates how the conception of social entities with emergent causal powers may be deployed to explain causal influences that have previously been attributed to rather amorphous conceptions of social structure, and to explain them in a rather less mystifying fashion than these earlier theories of structure. This has a number of important implications for the questions at issue here. First, it substantiates the notion that there may be social entities that have emergent causal powers, powers that would not exist in the absence of the entities concerned, and therefore casts doubt on methodological individualisms which argue that the idea of structural

---

<sup>11</sup> For further discussion of these and other complexities see Elder-Vass (2010: chapter 6).

power is really just a misdescription of the powers of individuals. Second, it provides a clear explanation of how such emergent causal powers may come about as the result of interactions between lower level parts of the social entities concerned, including (but not necessarily restricted to) human individuals. It thus provides a weak reduction of such powers while also providing grounds for denying that form of strong or eliminative reductionism we know as methodological individualism, since it describes a form of social causal power that depends on the existence of specific sorts of groups, and not just on the existence of human individuals.

### 3.5 Conclusion

This paper has argued that conceptions of emergence derived from the residual Cartesianism that has shaped the debate in the philosophy of mind have misled philosophers about what can usefully be achieved with emergence theory. This has led them to set a standard of unexplainability for emergence theories that is both unachievable and also inconsistent with our scientific understandings of the world. A more realistic and achievable objective for emergence theory, and one that is more consistent with scientific practice, is that it provides a justification for ascribing causal powers to entities while allowing that we may also be able to develop causal explanations of such powers. The form of relational emergence theory summarised in this paper does provide just such a justification and this is all that we need from emergence theory.

This is also, however, a form of emergence theory that has significant implications for the social sciences. It provides a clear justification for rejecting methodological individualism. But it also prompts an approach to social explanation in which we identify the social entities that are exercising causal powers, and the mechanisms through which they do so. Such an approach has the potential to bring explanatory clarity to a number of areas where it has been lacking in the past; this paper's discussion of norm circles as the social entities behind normative social institutions illustrates the point. It is an approach, therefore, that does not simply validate existing theories of social structure, but rather requires a new approach to theorising social structure, one that includes examination of the processes of interaction between individuals that generate social powers.

### References

- Allen, C., Todd, P. M., & Weinberg, J. M. (2012). Reasoning and rationality. In E. Margolis, R. Samuels, & S. P. Stich (Eds.), *The Oxford handbook of philosophy of cognitive science* (pp. 41–59). Oxford: Oxford University Press.

- Archer, M. S. (1995). *Realist social theory: The morphogenetic approach*. Cambridge: Cambridge University Press.
- Bedau, M. A. (1997). Weak emergence. *Philosophical Perspectives*, 11, 375–399.
- Bhaskar, R. (1975). *A realist theory of science*. Leeds: Leeds Books.
- Block, N. (2003). Do causal powers drain away? *Philosophy and Phenomenological Research*, LXVII, 133–150.
- Chernilo, D. (2007). *A social theory of the nation-state*. Abingdon: Routledge.
- Elder-Vass, D. (2005). Emergence and the realist account of cause. *Journal of Critical Realism*, 4, 315–338.
- Elder-Vass, D. (2010). *The causal power of social structures*. Cambridge: Cambridge University Press.
- Elder-Vass, D. (2011). The causal power of discourse. *Journal for the Theory of Social Behaviour*, 41, 143–160.
- Elder-Vass, D. (2012a). *The reality of social construction*. Cambridge: Cambridge University Press.
- Elder-Vass, D. (2012b). Top-down causation and social structures. *Interface Focus*, 2, 82–90.
- Gell-Mann, M. (1995). *The quark and the jaguar*. London: Abacus.
- Harré, R., & Bhaskar, R. (2001). How to change reality: Story v. structure – A debate. In J. López & G. Potter (Eds.), *After postmodernism: An introduction to critical realism* (pp. 22–39). London: Athlone Press.
- Kim, J. (1993). The non-reductivist's troubles with mental causation. In J. Heil & A. R. Mele (Eds.), *Mental causation* (pp. 189–210). Oxford: Clarendon.
- Kim, J. (1998). *Mind in a physical world*. Cambridge, MA: The MIT Press.
- Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, 95, 3–36.
- Kim, J. (2006). Emergence: Core ideas and issues. *Synthese*, 151, 547–559.
- Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Searle, J. R. (2002). Why I am not a property dualist. *Journal of Consciousness Studies*, 9, 57–64.
- Stephan, A. (2002). Emergentism, irreducibility, and downward causation. *Grazer Philosophische Studien*, 65, 77–94.
- Stones, R. (2005). *Structuration theory*. Basingstoke: Palgrave Macmillan.
- Varela, C. R., & Harre, R. (1996). Conflicting varieties of realism. *Journal for the Theory of Social Behaviour*, 26, 313–325.
- Wimsatt, W. C. (2000). Emergence as non-aggregativity and the biases of reductionisms. *Foundations of Science*, 5, 269–297.

## Chapter 4

# Actor-Centered Sociology and the New Pragmatism

Daniel Little

**Abstract** Theory and research in sociology need to be grounded in the fundamental truth that social phenomena are constituted by the actions and thoughts of socially constituted and socially situated individuals. This truism may be described as “methodological localism.” This does not imply that explanations must proceed from individual to social; but it does imply that we need to be confident that our hypotheses about social entities and processes have “microfoundations” at the level of the actors who constitute them. The article draws out an important consequence of this set of ideas: the necessity for sociology of developing a more adequate theory of the actor—an account of the ways the individual represents the world, the things that motivate him or her, and the ways that he or she arrives at actions and plans based on these features of practical cognition. To date the most common theory of the actor in the social sciences is the rational-intentional model and its cousin, rational choice theory. However, American pragmatism offers a significantly richer framework in terms of which to understand actors and their actions. This framework emphasizes habit, practice, and creativity in the genesis of action. Contemporary sociologists such as Neil Gross, Andrew Abbott, Mark Granovetter, and Hans Joas have taken this framework seriously in their theorizing with good effect. The article concludes that sociology gains when researchers arrive at more nuanced understandings of the constitutions and situations of the actors with whom they are concerned.

---

D. Little (✉)

Department of Philosophy, University of Michigan-Dearborn, Dearborn, MI, USA

Department of Sociology, University of Michigan-Ann Arb, Ann Arbor, MI, USA

e-mail: [delittle@umich.edu](mailto:delittle@umich.edu)

## 4.1 Introduction

I am attracted to an approach to sociological thinking that can be described as “actor-centered.” The basic idea is that social phenomena—political systems, organizations, institutions, social practices, normative systems, riots, cities, and mentalities—are constituted by the actions and thoughts of individuals, oriented by their own subjectivities, mental frameworks, and social relationships. It is recognized, of course, that the subjectivity of the actor does not come full-blown into his or her mind at adulthood; rather, individuals are “socialized” through a continuing series of experiences in family, school, playground, street, mosque, and social circles. Their thought processes and mental frameworks are developed through myriad social relationships and institutions. So the actor is a socially constituted individual.

This conception plays a key role in what I call a theory of methodological localism, in contrast to both methodological individualism and methodological holism (Little 2006, 2012a, b). Against classical methodological individualism, this view insists on the inherent “social-ness” of the actor who is both socially constituted and socially situated.<sup>1</sup> And against holism, it avoids the error of reification of large social forces disconnected from the individuals whose actions and mentalities embody them.

Methodological localism emphasizes two ways in which actors are socially embedded. In any given situation individuals are embedded within a set of social relations and institutions that create opportunities and costs for them. They have friends and enemies, they have bosses and workers, they have neighbors and co-religionists, they have families, they belong to political parties. All of these relations and institutions serve to constitute the environment within which they make plans and perform their actions. This complex setting of opportunities and regulative systems falls at the center of research for the new institutionalism (Brinton and Nee 1998; DiMaggio and Lincoln 1995; Ostrom 1990). And, as the new institutionalists rightly insist, there are deeply important variations across social space in the details of the workings of institutions and social networks. Two adjacent California counties may have slightly different rules of livestock liability; and these rule differences will lead to different patterns of behavior (Ellickson 1991).

The second form of social embeddedness is deeper and more persistent. The individual’s values, commitments, emotions, social ideals, repertoires of action, scripts of behavior, and ways of conceiving of the world are themselves the

---

<sup>1</sup> Advocates of analytical sociology favor a version of methodological individualism identified as “structural individualism”. “Structural individualism is a methodological doctrine according to which social facts should be explained as the intended or unintended outcomes of individuals’ actions. Structural individualism differs from traditional methodological individualism in attributing substantial explanatory importance to the social structures in which individuals are embedded” (Hedström and Bearman 2009: 4).

products of a lifetime of local social experiences. Individuals are socialized throughout their childhoods and adult lives into specific ways of thinking and acting, and the mosaic of these experiences serves to constitute the moral, emotional, and practical characteristics of the individual's social-cognitive system. The way the individual thinks about the social world is itself a feature of his/her social setting. Moreover, the mechanisms of socialization—schools, religious institutions, military experience, playgrounds, families—are themselves concrete social phenomena that are amenable to empirical sociological investigation, and they too are locally embodied. If we want to know why affluent Pakistani teenagers applauded on Facebook the murder of Punjab Governor Salmaan Taseer for his opposition to harsh blasphemy laws, then we need to look in detail at the ways in which the political and religious attitudes of this segment of Pakistani society took shape (Hanif 2011).

Essentially, then, the view comes down to this. The molecules of social life are socially constituted, socially situated individuals involved in proximate social relationships with other such individuals. The peasant in a village finds him- or herself in proximate relations to family members, neighbors, policemen, tax collectors, food sellers, and numerous other social actors; the employee in a corporation has proximate relations with family members, friends, co-workers, supervisors, subordinates, sub-contractors, and occasionally the CEO through an encouraging email. No one has proximate relations with the “national taxation system”; instead, they have communications from tax offices, visits from tax assessors, and peremptory demands from auditors when their tax papers do not add up. The national tax system does indeed influence them; but it does so through specific pathways of influence that eventually take the form of proximate contact. And the social characteristics of these various extended systems depend on the nature of these pathways and the actions that the various actors take along the way.

If our social ontology maintains that complex social processes and assemblages take shape out of the actions and thoughts of individuals, then it is logical that we need to develop a theory of the actor. This does not imply that our explanations always need to proceed from individual level to social level; sometimes this is an appropriate explanatory strategy and sometimes it is not. But even in circumstances where our explanations include hypotheses that refer to social entities, we still need to have some idea of the kinds of actions and interactions that establish the properties of those entities. We would like to have a justifiable set of ideas about how individuals perceive the social world, how they think about their own lives and commitments, what motivates them, and how they move from thought to action. But we have many alternatives available as we attempt to grapple with this task. Much social science theorizing depends on an over-simple theory of the actor, often involving the Aristotelian ideal of means-end rationality. A central thrust of the current paper is that we need to have a more adequate and nuanced theory of how individuals frame and guide their actions.

So what is an actor-centered sociology? It is an approach to sociological research and explanation that focuses our attention on the ways in which the actions and context of socially specified actors bring about various interesting sociological

outcomes. It permits the sociologist to hypothesize about social causes and influences at a variety of levels of aggregation; but it emphasizes the importance of being able to link our hypotheses about these higher-level social causes to the socially situated individuals who make them up (Little 2012b). The researcher is encouraged to keep in mind the central premise of methodological localism—the idea that all social forces, causes, and structures depend ultimately on the socially constituted and socially situated actors whose actions and thoughts make them up.<sup>2</sup>

Here is an elliptical description of three aspects of the idea of actor-centered sociology. First, it reflects a view of social ontology: Social things are composed, constituted, and propertied by the activities and interactions of individual actors—perhaps two, perhaps 300 million. Second, it puts forward a constraint on theorizing: Our social theories need to be compatible with the ontology. We need to be confident that there are individual-level mechanisms that underlie the causal properties and structures that we postulate. The way I put the point is this: it must be credible that our social theories, hypotheses, and assertions could be provided with microfoundations at the level of socially situated actors (Little 1994). Third, actor-centered sociology represents a heuristic about where to focus at least some of our research energy and attention: at the ordinary processes and relations through which social action take place, the ordinary people who bring them about, and the ordinary processes through which the effects of action and interaction aggregate to higher levels of social organization.

One issue that is somewhat troubling with this topic is that there seems to be an apparent inconsistency in two lines of thought that I have advocated for equally strongly: first, that social facts require microfoundations; and second, that meso-structures like organizations, institutions, and normative systems can have autonomous causal properties (Little 2012b). Are these two ideas consistent?

In particular, one might interpret the imperative of actor-centered sociology as a particularly restrictive view of social causation: from configurations of actors to meso-level social facts. So all the causal “action” is happening at the level of the actors, not the structures. This does not follow from the ontology of methodological localism adopted here. Dave Elder-Vass attempts to avoid this implication by arguing for emergent social causal properties (Elder-Vass 2010); I have approached the problem by talking about relatively autonomous causal properties at the meso-level (Little 2012b). I continue to think the latter view works reasonably well. In considering the institutions of the university as causal structures, for example, I think a plausible case is made for both ideas: the tenure system is causally effective in constraining individual faculty members’ behavior as well as being causally effective in influencing other structural features of the university; **and** we can be confident that every aspect of this system can be provided with microfoundations in

---

<sup>2</sup> Sociologists in the tradition of James Coleman also pursue an “actor-centered” sociology (Coleman 1990); but their approach is grounded in rational choice theory and a more rigidly methodological-individualist perspective. An actor-centered approach that embraces the methodological localism perspective is more accepting of the inter-penetration of individual action and social influences, and has an ontology that justifies this acceptance.

the form of the structured circumstances of action and cultururation through which the bureaucratic agents in the system behave. Or in other words: it is consistent to maintain both parts of the dilemma, actor-centered sociology and relatively autonomous meso-level social causation.

## 4.2 Theories of the Actor

This discussion suggests that some of the fundamental questions for the social sciences are these: Who are the actors? How do the actors interact with each other? How do these actions and interactions aggregate to larger structures and processes?

We think we know quite a bit about being an “actor”. It is to be an intervener in the world: a being capable of changing things around him or her through physical action; a subjectivity (consciousness, feelings, thoughts, desires); a user of language for thought and for communication. We also think we are well prepared when it comes to recognizing the diversity of the subjectivities on which agents act and live. We are familiar with important cultural differences; we know that people have different sets of values and commitments; we know that people from different backgrounds see and experience the world differently. Perhaps even, like Benjamin Whorf, we may think that different language systems and grammars give their users fundamentally different categories in term of which to analyze the world (Whorf 1956).

So we seem to know a lot. But in fact, this is the part of the social sciences and humanities that is the least developed and the least adequate to the complexity of the facts. Moreover, the assumptions commonly made about actors and their actions are often superficial and misleading. Individual actors’ behavior is not as simple as rational choice theory would have it, and it is helpful to look more closely at the ensemble of habits, learning, spontaneity, emotion, and routine that go into everyday behavior. This deeper knowledge of the ways that actors think and act is likely to be productive for social scientists at every level, from the ethnographic sociologist to the organizational behavior specialist to the scholar of the emergence of fascist politics and mobilization. And sometimes this knowledge will contribute to a particular kind of explanatory success: an individual-level mechanism producing a meso-level effect.

The topic is important, because ultimately, all social phenomena are the result of agents acting for their own reasons, and we need to understand the varieties of their thought processes. This is directly true when we seek to explain a social outcome as a result of the actions of individuals. But it is also true when we explain one social outcome on the basis of other social circumstances and structures. We still need to have an idea of how human beings behaving as they do give rise to the social features of the social features cited in the explanation. Consider the hypothesis that some organizations have the capacity to learn and adjust according to past experience. This poses a key question: How could an organization act as though it had an intelligent system for processing its experience? If we have some concrete idea of



how individuals in organizations behave; if we have some idea of how the communications processes and transmissions occur; if we have some idea of how supervisors bend the activities of their subordinates to the needs of the organization; then we are likely to have a good idea about what some of the plausible mechanisms are that might transmit the organization from experience to adaptive behavior.

It would appear, then, that understanding some of the variety of ways that real human beings think, act, and behave is genuinely important for most social-science research questions. If we believe ontologically that social structures, processes, causal relations, patterns, etc., are all ultimately constituted or created or caused by the actions and interactions of individual actors who make them up, then it seems plausible that we are better off knowing more about their dispositions, mental systems, decision processes, etc. Moreover, a better understanding of the diversity of human agency may offer indications of what sorts of mechanisms might be in play in specific puzzling social circumstances. A richer theory of actor may help to guide social science research towards more fruitful hypotheses about underlying mechanisms. Thus we have both ontological and heuristic reasons for thinking that a better conception of the actor will be productive for the social sciences. It is useful for cell biologists to know quite a bit about biochemistry, even though their explanations remain at the cellular level.

So the social sciences will be well served by developing more detailed understandings of the varieties of human subjectivity and agency.

### 4.3 What Does a Theory of the Actor Do?

We might begin by asking what we want from a theory of the actor. A simple answer is this: a theory of the actor ought to provide a basis for analyzing the mental processes of the real human actor as he or she proceeds through life activity. One individual decides to stop by a retirement home to visit an elderly friend; another individual breaks into a car to steal a briefcase; another has an argument with her boss and decides to quit her job. What sorts of thinking go into these choices and innumerable others?

These are all actions that are to some extent deliberate and considered; and yet they are also complicated responses to shifting circumstances and events that have prudential and emotional meanings to the actors in question. Here I am not asking the question, "What sorts of factors motivate people, and how do they process their motivations?", though these are interesting and important questions. Rather, I am asking a more basic question: what categories and concepts should we use in attempting to analyze deliberate action in the first place? What features of mental experience need to be highlighted? Should we expect there to be a best framework of analysis of the components of practical mentality? Or is this fated always to be an open question?

Think of the range of vocabulary that is relevant to our discourse about the kinds of examples mentioned above: decision, belief, desire, emotion, fear, habit, norm,

obligation, reason, impulse, weakness of will . . . These terms and others constitute something like a mental ontology, a set of concepts that we attribute to the agent as he/she acts. And some of them bring presuppositions that are debatable. Take “decision,” for example. Did the thief “decide” to break into the car? Or was the action an impulse, prior to thought and deliberation?

Aristotle’s writings guide much philosophical thinking on these questions by offering an orderly theory of the practical agent. His theory is centered on the idea of deliberative rationality, but he leaves a place for the emotions in action as well (to be controlled by the faculty of reason). Deliberation, in Aristotle’s view, amounts to reflecting on one’s goals and arranging them into a hierarchy; then choosing actions that permit the achievement of one’s highest goals.<sup>3</sup> Rational choice theory provides a contemporary interpretation of this Aristotelian approach. Actors have preferences and beliefs; their preferences are well ordered; they assign probabilities and utilities to outcomes (the results of actions); and they choose a given action to maximize the satisfaction of their preferences.

Here is the crucial point, however: simplified models of means-end rationality do not do justice to most instances of real social action and interaction. We need to develop more nuanced understandings of the actor and agency if we are to provide a convincing basis for sociological explanation.

#### 4.4 Pragmatist Theory

There are alternatives to the basic Aristotelian model of goal-directed agency. The research of Erving Goffman (1956) and Harold Garfinkel (1967) offered one possible alternative—highly detailed efforts at discovering the micro-level details of human social behavior, or what Garfinkel referred to as ethnomethodology. Goffman, Garfinkel, and other micro-sociologists shared a commitment to careful, detailed observation and description of social behavior. They were interested in capturing the nuances of ordinary behavior, and their research reports give a great deal of emphasis to the importance of providing detailed descriptions of ordinary social interactions.

So Goffman started much of his work with fine-grained, detailed observations of behavior—elements as subtle as the bodily signs of embarrassment in a conversation (Goffman 1967). Second, he wanted to discover some of the structures of mental processing through which individuals act in social settings—the frames, scripts, and rituals that guide social perception and action for individuals. Third, he wanted to tease out the social judgments and norms that provide the normative and guiding context for the actor’s movements—the judgments surrounding “saving face,” embarrassment, shame, pride, and proper performance. This constitutes an

---

<sup>3</sup> Christine Korsgaard’s writings about practical reason shed light on Aristotle’s assumptions about agency (Korsgaard 2008).

important contribution to the topic at hand—the problem of formulating a more adequate theory of the actor.

Several other important alternatives to the Aristotelian model of agency have developed through the work of Pierre Bourdieu (1977), Anthony Giddens (1979), and Margaret Archer (2000). Bourdieu's theory of practice and *habitus* emphasizes the spontaneity of human action, even while it allows for purposiveness and intentionality. Here is Bourdieu's description of *habitus*: “the durably installed generative principle of regulated improvisations, produces practices which tend to reproduce the regularities immanent in the objective conditions of the production of their generative principle, while adjusting to the demands inscribed as objective potentialities in the situation, as defined by the cognitive and motivating structures making up the *habitus*” (Bourdieu 1977: 78).

What this definition conveys is a more fluid conception of thinking-experiencing-doing on Bourdieu's part than is characteristic of a more Aristotelian conception. There is a suggestion of a sort of tacit knowledge underlying activity, with action looking as much like the smooth, intelligent motions of a soccer player as a deliberative chess master. It is a less epistemic view of the human condition—less a concoction of explicit beliefs and reasonings than a smooth coordination of tacit understandings and movements in the situation. It is less about deliberation and decision and more about intelligent doing.

Anthony Giddens offered an influential theory of “structuration” as an account of how actions take shape within a social context. His view is that the two poles of structure and agency must be considered from within a common formulation. Like Bourdieu, Giddens also emphasizes the idea that actions take place in a stream of thinking and doing: “Action or agency, as I use it, thus does not refer to a series of discrete acts combined together, but to a continuous flow of conduct. We may define action . . . as involving a ‘stream of actual or contemplated causal interventions of corporeal beings in the ongoing process of events-in-the-world’” (Giddens 1979: 55).

Margaret Archer too takes issue with the standard views of agency and structure. She proceeds within the general framework of critical realism associated with Roy Bhaskar (1975). She refers to her view as the “morphogenetic” approach. Here is how she explains this concept.

The ‘morpho’ element is an acknowledgement that society has no pre-set form or preferred state: the ‘genetic’ part is a recognition that it takes its shape from, and is formed by, agents, originating from the intended and unintended consequences of their activities. (5)

These are avenues worth pursuing today. What they lack, however, is a sustained effort to provide a nuanced theory of the actor. But a more comprehensive framework for thinking about alternative theories of the actor has begun to emerge in new thinking about the social world inspired by American pragmatism. This includes work by such sociologists as Neil Gross, Andrew Abbott, Mark Granovetter, and Hans Joas. This approach offers a rather different view of the action/actor nexus. Actors are considered to be less deliberative and rational than the Aristotelian version of action theory, and the ideas of habit, improvisation, and creativity are

given more prominence. Moreover, there is a shift of emphasis from the “actor” to the “action in context”, where the actor him/herself is driven by the relationships and the flow of action and need.

In particular, here are several provocative ideas advanced in the new pragmatism—

- “action is relational rather than individual”
- “action is often the manifestation of socially created ‘habit’”
- “action is a flow of improvisational adaptations”
- “focus on the action rather than the actor”

Relationality is the idea that the action cannot be understood unless it is placed into a field of related actions and actors. The idea of habit is the notion that actors develop repertoires of routine responses to life situations. Improvisation captures the idea of creativity and responsiveness in action. And focusing on the action rather than the actor de-dramatizes the Aristotelian notion of actors designing and implementing actions. Andrew Abbott captures many of these ideas in his recent work (Abbott 2007: 8).

The founding pragmatists had specific ideas about each of these topics. George Herbert Mead is one of the thinkers whose ideas have significantly influenced the current group of pragmatist sociologists. Mead’s theory postulates that the self is built up out of imitative practices, gestures, and conversations over time. The individual forms a reflective conception of his/her self that derives from example and engagement with specific other actors within his/her social space. Mead’s fundamental view is that the tradition of Enlightenment philosophy has gotten the relationship between individual and society backwards; philosophers have built the social from the individual, but actually the self is in some important way the sum of its social relations (Mead and Morris 1962: 222). Mead favors the “social first” approach. This does not rest on some kind of spooky Durkheimianism about irreducible social wholes, but rather the point that individuals always take shape within the ambit of a set of social relationships, language practices, and normative cues. (His view thus parallels a central idea in the doctrine of methodological localism, the idea that the individual is socially constituted.)

These ideas provide a very different framework for thinking about the actor as he or she engages in action and interaction in the world. It is a framework that places the formation of the actor’s ideas, beliefs, and motives squarely within the ongoing social environment within which she takes shape; and it understands the actions in which the actor engages as an integrated set of interactions with the persons with whom she is engaging.

Another major influence on contemporary pragmatist sociologists is John Dewey. Dewey’s analysis of “habit” is another linchpin of the pragmatist theory of action as reconstructed by the current generation of sociologists. Dewey believes that a large volume of our ordinary human conduct is not deliberative at all, but is rather based on habit. In the Aristotelian and Kantian tradition of deliberative rationality, the idea of will is central. The agent deliberates about ends and means and chooses (wills) a means that will bring about her ends. So the will is the

fundamental element of action. But in fact, Dewey argues that the idea of “will” itself can be understood as a compound of habits, rather than a self-originating deliberation about ends and means.

Dewey also emphasizes the point that habits in action generally presuppose a social context (Dewey 1922: 16). We acquire our habits of behavior through exposure to other actors. Both Mead and Dewey emphasize a point that was made earlier in the context of the idea of methodological localism: the individual takes shape through the persistent fact of existing social practices and norms. So, according to Dewey, much of our practical life—the domain in which we act and choose—is the result of an assembly of socially instilled habits, not rationally developed plans.<sup>4</sup>

So the fact of habit in action is fundamental to Dewey’s view of the social world and the individual actor’s activities within that world. And it is a role that suggests that Dewey differs very fundamentally from the Aristotelian view of deliberative rationality in action, where the actor identifies a set of ends, arrives at a set of beliefs, and reasons to a conclusion about what action to choose.<sup>5</sup> Dewey does not say that there is nothing deliberative about action, but he appears to believe that habit is more common and more fundamental. Further, he seems to believe that many examples of the exercise of will are in fact examples of the influence of nested sets of habits.

## 4.5 The New Pragmatism

Now let us see how contemporary sociologists have developed some new versions of a pragmatist theory of the actor. What might a pragmatist theory of action involve? This approach takes issue with models of deliberative rationality because—as Andrew Abbott and Neil Gross maintain—they privilege the actor over the action, the individual over the interaction. They push us in the direction of a social ontology that is individualistic and perhaps reductionist. Abbott proposes, in contrast, that we begin with the interaction, the flow of moves and responses. Gross suggests starting with the creativity inherent in any complex flow of human activities and interactions. And Granovetter emphasizes the insufficiency of the customary distinction between rational and normative behavior in explaining action. This part of the paper concludes with a review of the most systematic current effort to provide an alternative pragmatist action theory, which is offered by Hans Joas.

---

<sup>4</sup> Erkki Kilpinen offers a careful analysis of the role of the pragmatist conception of habit in sociology (Kilpinen 2009).

<sup>5</sup> The idea of action as habit seems to have more in common with another aspect of Aristotle’s theory of action, the role that virtue plays in ordinary conduct (Korsgaard 2008).

### 4.5.1 *Neil Gross*

Neil Gross makes use of many of these insights from pragmatism in his efforts to formulate a more adequate framework of sociological thinking (Gross 2009). Gross's work attempts to improve upon the theory of action that much social theory has presupposed for more than a century—the idea of the rational, purposive agent considering options and choosing outcomes. Against that hyper-deliberative conception, Gross (and pragmatism) advocates for a more fluid, interactional, and only partially conscious flow of actions. Here is how Gross characterizes the import of pragmatism for sociology.

Drawing on the work of Hans Joas (1993, 1996), I argued that at the core of the tradition lies a conception of human action from which, for the classical pragmatists, controversial epistemological implications followed. This conception holds that human beings act in the service of solving practical problems they confront in the course of their daily lives, and that their action takes the form of an alternation between more or less habituated patterns of response and creative improvisation and experimentation when habit proves inadequate. (Gross 2010: 338)

This passage emphasizes habit, improvisation, and creativity rather than deliberation and choice. There is a suggestion here of stylized modes of behavior (scripts) within which persons locate their actions, and a suggestion of the importance of specific cognitive fields embodied in social groups that contextualize and rationalize the person's activities (assumptions, for example, of how a doctor should treat a patient in a hospital).

Another important part of Gross's conception of pragmatist action theory is the way we conceive of the individual. According to the pragmatist theory, the individual needs to be considered within the context of a social group, influenced by norms, emotions, and actions of the others in the group. So action should not be "atomized" into a group of individual actors choosing independently.

Third, there is the question of how the agent decides what to do in a particular circumstance. The pragmatist view that Gross describes holds that the actor chooses in line with habit and script. Essentially, this is the insight that there are fairly well defined rules of thumb or scripts for how to respond to certain kinds of problems. And the theory holds that the actor generally acts accordingly. When an experienced politician is confronted by a heckler, the play book pretty well specifies how he/she should respond. This contrasts sharply from the deliberativist view of action. So habit and improvisation are key to action, rather than rational planning and calculation about means and ends.<sup>6</sup>

What makes this set of assumptions about action a "pragmatist" approach? Fundamentally, because it understands the actor as situated within a field of assumptions, modes of behavior, ways of perceiving; and as being stimulated to action by "problem situations". So action is understood as the actor's creative use of

---

<sup>6</sup>Gross puts this approach into play in his extended sociological biography of Richard Rorty (Gross 2008).

scripts, habits, and cognitive frameworks to solve particular problems. (Gross refers to this as an A-P-H-R chain: actor, problem situation, habit, and response; Gross 2010: 343.)

This framework has important implications about the actor's state of mind in performing an action. It forces us to consider in new ways the degree to which the actor's own agency determines his/her actions. To what extent is the actor's conduct his own, deriving from his or her own authentic self rather than from an externally created role? Consider as an example, a fire captain's behaviors and speeches during his management of the fire company's handling of a major fire. The captain and the members of the company have a well-rehearsed set of procedures for various scenarios encountered in fighting a fire—"victim trapped in bedroom", "elderly person inside", "possible toxic materials on site", "roof collapse imminent". As these contingencies arise, captain and company play out the prescribed actions. Second, though, there is probably a prescribed style of command that influences the captain's manner and conduct: be calm, give sufficient attention to the company's safety, keep control of the press and crowd on-site, don't fall into a screaming, cursing rage when things go badly.

How does this conception of action compare to the foil of pure deliberative rationality? According to rational choice theory an actor makes a choice in a problem situation by (i) arranging a preference ordering of possible outcomes, including utilities for each outcome; (ii) consulting rational procedures to gain beliefs about the probabilities of various strategies leading to various outcomes; and (iii) choosing that strategy that results in the greatest expected utility (utility  $\times$  probability). This account makes choice rational in both aspects: rational acquisition of beliefs about interventions and outcomes, and rational comparison of the relative goodness/badness of the outcomes associated with possible interventions. There is no place in this rational-choice story for culturally variable cognitive frameworks for perceiving the situation, or for group-specific rules of thumb governing the choice of interventions.

In short, Gross argues that the pragmatist approach is a fertile one for researchers interested in analyzing and explaining complex processes of social action. His own work demonstrates the fertility of pragmatist ideas about agency. His sociological biography of Richard Rorty is a good example (Gross 2008). Here he offers a sociology of ideas that brings agency back in. He introduces the idea of the role of the individual's "self-concept", which turns out to be a basis for the choices the young intellectual makes within the context of the strategy-setting realities of the field. A self-concept is a set of values, purposes, and conceptions that the individual has acquired through a variety of social structures, and that continues to evolve through life. Gross emphasizes the narrative character of a self-conception: it is expressed and embodied through the stories the individual tells him/herself and others about the development of his/her life. And Gross thinks that these stories are deeply influential, in terms of the choices that a developing intellectual makes at each stage of life. In particular, he thinks that the academic's choices are often inflected by his/her self-conception to an extent that may override the strategic and prudential considerations that are highlighted by Bourdieu (1996) and Collins

(1998) in their treatments of academic careers. In his view, identity trumps interest—at least sometimes.

### 4.5.2 *Andrew Abbott*

Another important contemporary sociologist who has been influenced by a pragmatist theory of the actor is Andrew Abbott. Abbott explicitly ties his approach to that of pragmatism: “In this paper I shall consider the mechanisms movement from the viewpoint of a different theoretical tradition, one that focuses on the processual and relational character of social life and that traces its roots to pragmatism” (Abbott 2007: 2).<sup>7</sup> Interaction and social relations are key to the theory of action that Abbott proposes. Abbott takes issue with the project of treating the actor as a well defined unit. Instead he offers arguments for the primacy of a relational approach and actions-in-context instead of unitary actors. Much of this line of thought comes down to Abbott’s view that actors and agency are deeply socially constructed; so it does not make sense to take the actor as a given who then deliberates about options.

Making interaction primitive makes it possible to give an account of the self. By making the self be continuously recreated in the flow of interaction we bring it out of the realm of assumptions and into that of investigation. At the same time, by making interaction primitive we allow for the endless interplay of cross-individual structural definitions of the flow of action, an interplay that is an evident fact in social life. (Abbott 2007: 8)

So Abbott too rejects the idea of a self-sufficient, deliberative actor who surveys a range of choices and chooses an action to further his or her purposes. Instead, he argues that action is substantially more improvisational and habitual than this picture allows. The actor is created by her actions, rather than the author of those actions (12).

Abbott turns these criticisms against some of the key premises of analytical sociology, because he believes that the analytical sociologists have too uncritically adopted a rational-choice model of the actor (Abbott 2007).<sup>8</sup> In particular, Abbott argues that the idea of a social mechanism is impaired by its dependency on methodological individualism.<sup>9</sup> He believes that the mechanisms approach has tended towards reductionism to the level of the rational individual; he would prefer a theory of mechanisms that rests on a relational theory of action.

---

<sup>7</sup> The mechanisms movement is the view advocated by Peter Hedström, Jon Elster, and others that social explanations need to be couched in terms of the social causal mechanisms that bring about structures and outcomes of interest (Hedström 2005), (Hedström and Swedberg 1998), (Elster 2007).

<sup>8</sup> Peter Hedström’s *Dissecting the Social* provides a key statement of the approach to sociology taken by analytical sociology (Hedström 2005).

<sup>9</sup> For more on this debate see “Analytical Sociology and the Rest of Sociology” (Little 2012a).



### 4.5.3 *Mark Granovetter*

Mark Granovetter is another sociologist who has cast doubt on the standard rational-intentional model of the actor. Granovetter formulated a foundational set of questions for the social sciences in “Economic Action and Social Structure: The Problem of Embeddedness” (Granovetter 1985). To what extent do individuals choose their courses of action largely on the basis of a calculation of costs and benefits? And to what extent, on the contrary, are their actions importantly driven by the normative assumptions they share with other individuals with whom they interact? He used the concept of *embeddedness* as a way of capturing the idea that the actions individuals choose are importantly refracted by the social relations within which they function.

The large distinction at issue here is the contrast between rational actor models of the social world, in which the actor makes choices within a thin set of context-independent decision rules, and social actor models, in which the actor is largely driven by a context-defined set of scripts as he/she makes choices. Granovetter argues that neither of the polar positions is tenable. The formalist approach errs in taking too a-social view of the actor. But the extreme norm-driven alternative isn’t appealing either.

So action does not reduce to abstract optimizing rationality, and it does not reduce to inflexible cultural or normative scripts either. Instead, Granovetter proposes an approach to this topic that reframes the issue around a more fluid and relational conception of the actor. Like the pragmatist theories of the actor associated with Abbott and Gross, he explores the idea that the actor’s choices emerge from a flow of interactions and shifting relations with others. The actor is not an atomized agent, but rather a participant in a flow of actions and interactions.

At the same time, Granovetter insists that this approach does not deny purposiveness and agency to the actor. The actor reacts and responds to the social relations surrounding him or her; but actions are constructed and refracted through the consciousness, beliefs, and purposes of the individual.

The idea of an “embedded” individual is contrasted to the idea of an atomized actor; this implies that the individual’s choices and actions are generated, in part anyway, by the actions and expected behavior of other actors. It is a relational concept; the embedded actor exists in a set of relationships with other actors whose choices affect his or her own choices as well. And this in turn implies that the choices actors make are not wholly determined by facts internal to their spheres of individual deliberation and beliefs; instead, actions are importantly influenced by the observed and expected behavior of others. Granovetter makes use of the idea of “embeddedness” to capture this feature of action: “Their attempts at purposive action are instead embedded in concrete, ongoing systems of social relations” (487).

Some of Granovetter’s discussion crystallizes around the social reality of trust within a system of economic actors. Trust is an inherently relational social category; it depends upon the past and present actions and interactions within a group of

actors, on the basis of which the actors choose courses of action that depend on expectations about the future cooperative actions of the other actors. Trust for Granovetter is therefore a feature of social relations and social networks. And trust is relevant to cooperation in all its variants—benevolent and malicious as well. As Granovetter points out, a conspiracy to defraud a business requires a group of trusting confederates. So it is an important sociological question to investigate how those bonds of trust among thieves are created and sustained.

This line of thought, and the theory of the actor that it suggests, is an important contribution to how we can understand social behavior in a wide range of contexts. The key insight is that individuals choose their actions in consideration of the likely choices of others, and this means that their concrete social relations are critical to their actions. How frequently do a set of actors interact? Has there been a history of successful cooperation among these actors in the past? Are there rivalries among the actors that might work to reduce trust? These are all situational and historical facts about the location and social relations of the individual. And they imply that very similar individuals, confronting very similar circumstances of choice, may arrive at very different patterns of social action dependent on their histories of interaction with each other.

#### 4.5.4 *Hans Joas*

The German sociologist Hans Joas has gone a long ways towards providing an explicit formulation of a pragmatist alternative for the theory of action. Joas has written extensively on G. H. Mead, and has also made substantial efforts to articulate a theory of action that corresponds to some of the core ideas of these strands of American pragmatism (Joas 1985, 1993, 1996; Joas and Beckert 2001). Here is how Joas puts his view of the field in *The Creativity of Action*: “The central thesis in this book is the claim that a third model of action should be added to the two predominant models of action, namely *rational* action and *normatively oriented* action. What I have in mind is a model that emphasizes the *creative* character of human action” (Joas 1996: 4).

So what is a “creative” theory of action? Joas’s article with Jens Beckert on “Action Theory” in Jonathan Turner’s *Handbook of Sociological Theory* is a good place to start (Joas and Beckert 2001), since Joas and Beckert are specifically concerned there to give an exposition of a theory of action that is grounded in a pragmatist account.

Joas and Beckert begin their account by framing the standard assumptions of existing action theory in terms of two poles: action as rational choice (e.g. James Coleman) and action as conformance to a set of prescriptions and norms (e.g. Durkheim, Parsons). (Note the parallel here with Granovetter’s starting point.) They argue for a view that is separate from both of these, under the heading of “creative action”.

Common to both traditional views is the assumption of purposiveness: that action proceeds to bring about explicit pre-articulated goals subject to antecedently recognized constraints. The pragmatist view of action rejects this separation between goals, action, and outcome, and focuses on the fact that goals and actions themselves are formulated within a dynamic and extended process of thought and movement.<sup>10</sup> Tactics, movements, and responses are creative adaptations to fluidly changing circumstances. The basketball player driving to the basket is looking to score a goal or find an open teammate. But it is the rapid flow of movement, response by other players, and position on the floor that shapes the extended action of “driving for a layup.” Likewise, a talented public speaker approaches the podium with a few goals and ideas for the speech. But the actual flow of ideas, words, gestures, and flourishes is the result of the thinking speaker interacting dynamically with the audience. Joas and Beckert put their view in these terms:

At the beginning of an action process goals are frequently unspecific and only vaguely understood. They become clearer once the actor has a better understanding of the possible means to achieve the ends; even new goals will arise on the basis of newly available means. . . . For the theory of creativity of action the significance of the situation is far greater: Action is not only contingent on the structure of the situation but the situation is constitutive of action. (273, 274)

So what are the features of the situation that intersect with the thinking actor to create the temporally extended action? Joas and Beckert refer to corporality and sociality. The body is not simply the instrument of the agent. Rather, the physical features and limitations of the body themselves contribute to the unfolding of the action. (This aspect of the theory has much to do with phenomenology.) And the other persons involved in an action are not simply subjects of manipulation. Their own creativity in movement and action defines the changing parameters of the actor’s course of action. (Again, think of the analogy of 10 players in a basketball game.) Joas and Beckert think that this interpretation of action as extended intelligent adaptation to shifting circumstances helps to account for complex social circumstances that rational-actor and normative-actor theories have difficulty with.

This is a rich and nuanced theory of action, and one that has the potential for offering a basis of a much richer analysis of concrete social circumstances than we currently have. At the same time it should not be thought to be in contradiction to either rational-deliberation or normative-deliberation theories. These creative actors whom Joas describes are purposive in a more diffuse sense, and they are responsive to norms in action. It seems to me that the chief tension Joas offers is between stylized, mono-stranded models of action, and thick theories that incorporate the plain fact of intelligent adaptation and shaping of behavior that occurs in virtually all human activities.

---

<sup>10</sup> As demonstrated above, Dewey is the chief source of this view.

## 4.6 Import for Sociology

So far I have argued that pragmatism offers new resources for assembling a more adequate theory of the actor. Is this beneficial for the tasks that sociologists set themselves? In what ways can a more developed theory of the actor prove to be of value for empirical and theoretical work in the social sciences? The best way to provide support for a methodological innovation is to show how it informs successful new research. There are many strong contemporary examples of successful and innovative research that embody a more nuanced theory of the actor. I will close by considering one example in detail, George Steinmetz's careful attention to the causes of differentiation within the colonial locations of nineteenth-century Germany in *The Devil's Handwriting* (Steinmetz 2007).

*The Devil's Handwriting* offers a comparative historical sociology of German colonialism. Steinmetz wants to explain differences in the implementation of "native policy" within German colonial regimes around the turn of the twentieth century. He finds that there were significant differences across three major instances of German colonialism (Samoa, Qingdao, Southwest Africa), and he wants to know why. (For example, the Namibia regime was much more violent than the Samoa or Qingdao examples.) The reason this research is interesting in the current context is that Steinmetz makes use of an actor-centered approach that has a great deal in common with the ideas being discussed within the new pragmatism. So let us look at some of the details of the explanations that Steinmetz offers.

Here are the guiding explanatory ideas in Steinmetz's study.

What I try to account for in this book—my "explanandum"—is colonial native policy. Four determining structures or causal mechanisms were especially important in each of these colonies: (1) precolonial ethnographic discourses or representations, (2) symbolic competition among colonial officials for recognition of their superior ethnographic acuity, (3) colonizers' cross-identification with imagos of the colonized, and (4) responses by the colonized, including resistance, collaboration, and everything in between. Two other mechanisms influenced colonial native policy to varying degrees: [5] "economic" dynamics related to capitalist profit seeking (plantation agriculture, mining, trade, and smaller-scale forms of business) and [6] the "political pressures generated by the international system of states. (2)

The most important thing to note here is that all of the first four factors have to do with features of the actors—administrators, the colonized, the public. The thrust of Steinmetz's empirical work is to determine the specific circumstances that led to actors and actions of the kinds that produced the various colonialisms under study. Steinmetz suggests that all six factors mentioned here are causally relevant, but that the general structural causes (5 and 6) do not account for the variation in the cases. These structural factors perhaps account for the similarities rather than the differences across the cases. So it is differences in the circumstances and formation of the actors that provide the crux of Steinmetz's explanation.

Steinmetz provides support and explication for each of these four key factors. Consider first factor (2) above, the idea that the specifics of colonial rule depended a great deal on the circumstances of the professional and ideological "field" within

which colonial administrators were recruited and served. This idea makes use of Bourdieu's theory of the field (Bourdieu 1996). The idea here is that the particular intellectual and professional environment in Germany established certain points of difference around which participants competed. These dividing lines set the terms of professional competition, and prospective colonial administrators as well as functioning administrators needed to establish their program for governance around a distinctive package of these assumptions.

This mechanism provides a promising basis for explaining some of the otherwise puzzling aspects of colonial rule and native policy. Actors (administrators or other leaders) are immersed in a policy environment in which conflicting ideas about success are debated; the actors seek to align their actions to the framework they judge to be most likely to prevail (and further their careers).

Steinmetz also believes that the "ethnographic" discourses of the early twentieth century in Germany (factor 1) created the space of the imaginable when it came to colonial policy and strategy. The mentality, background assumptions, and value structures of the budding administrators were developed within the specific circumstances of the ideas and assumptions of late nineteenth-century Germany about the other—African, Samoan, Chinese. The explanation postulates a social fact—the prevalence of several intellectual frameworks about the other, several ethnographic discourses; and the individual's immersion in these discourses leads him or her to act intelligently and creatively in pursuing specific career goals.

This illustrates the way that factors (1) and (2) work in Steinmetz's explanation. What about (3)? This falls in the category of what Steinmetz calls "symbolic and imaginary identifications" (55). Here Steinmetz turns away from conscious calculation and mental representation towards aspects of a non-rational psychology. Here Steinmetz draws on psychoanalytic theory and the theories of Lacan. This part of the analysis makes use of assumptions about unconscious motivations and representations—a far cry from the rational, deliberative Aristotelian paradigm. But it remains an actor-centered analysis.

The fourth factor in Steinmetz's analysis involves the states of agency of the colonized. Here he refers to strategies of response by the subject people to the facts of colonial rule, ranging from cooperation to resistance. This aspect of the story too is highly compatible with the actor-center approach; it is straightforward to see how social mobilization theory can be fleshed out along the lines of an actor-centered approach.

So Steinmetz's account has several important characteristics. First, Steinmetz is interested in providing a contextualized explanation of differences in nominally similar outcomes (different instances of German colonial rule). Second, he is interested in providing an account of the social mechanisms that shaped each of the instances, in such a way as to account for their differences. Third, the mechanisms that he highlights are largely actor-centered mechanisms; they turn on the states of mentality, volition, and thinking of various of the participant actors. Fourth, the account deliberately highlights the contingency of the developments it describes. Individuals and particular institutions play a role, as well as historical occurrences that were themselves highly contingent. This feature of the account is

consistent with the creativity and spontaneity associated with the pragmatist theory of action. And finally, there is a pervasive use of social concepts like field, ideology, and worldview that play a crucial causal role in the story. These are factors that shape and inform the states of agency we observe in historically situated actors.

This example falls within the field of comparative historical sociology. But equally strong cases can be drawn from other areas of contemporary sociology, including the sociology of knowledge, urban sociology, criminology, and the sociology of political movements. Moreover, much of what is most satisfying and convincing about Steinmetz's case stems directly from the effective job he does in showing how these circumstances and conditions of the actor's mentality lead to details of difference in the resulting forms of colonial rule. So the focus on the actor pays off in substantive ways when it comes to arriving at empirically well grounded explanations of otherwise puzzling outcomes.

This case also validates several of the key imperatives of the doctrine of methodological localism articulated above: the insistence on the centrality of the socially situated and socially constructed actor within more complex social processes. Methodological localism implies that we need to be cautious about oversimplifying the mentality of the actor—not simply a utility maximizing egoist, not simply a norm-driven robot, not simply an adherent of a religious worldview. Instead, we need to pay attention to the details and the differences that we find in the historical setting of important social processes and outcomes: the specific forms of education received by scientists, the specific social environment in which prospective administrators were socialized, the specific mental frameworks associated with this or that historically situated community. These details help us to do a much better job of understanding how the actors perceived social situations and how they chose to act within them. And likewise, we need to pay attention to the regulative and incentive-generating context within which actors constructed their actions. This is the role that the intellectual and policy field plays in Steinmetz's account; it is also the role that specific property and contract arrangements play in the new institutionalism. And both Bourdieu and the new institutionalists are right that small differences in the institutional setting can result in large differences in outcome, as actors respond to institutions and incentives to pursue their ends. So paying close and detailed attention to the particulars of the institutions of career, economic opportunity, family, power, and prestige allows us to perceive the causes of important differences in outcomes.

In short, it seems that sociology has everything to gain by paying more attention to the specifics of the actors whose thinking and actions constitute the social processes of interest to them. This advice does not imply reductionism; it is entirely legitimate for sociologists to make use of causal claims at a variety of levels. But it does imply that there is substantive and valuable work to be done in almost every field of sociology at the level of the actor. Sociology gains when researchers attempt to gain a more nuanced understanding of the constitutions and situations of the actors with whom they are concerned.

## References

- Abbott, A. (2007). Mechanisms and relations. *Sociologica*, 2, 1–22. doi:10.2383/24750
- Archer, M. S. (2000). *Being human: The problem of agency*. Cambridge/New York: Cambridge University Press.
- Bhaskar, R. (1975). *A realist theory of science*. Leeds: Leeds Books.
- Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge: Cambridge University Press.
- Bourdieu, P. (1996). *The rules of art: Genesis and structure of the literary field* (Crossing aesthetics). Stanford: Stanford University Press.
- Brinton, M. C., & Nee, V. (1998). *New institutionalism in sociology*. New York: Russell Sage.
- Coleman, J. S. (1990). *Foundations of social theory*. Cambridge: Harvard University Press.
- Collins, R. (1998). *The sociology of philosophies: A global theory of intellectual change*. Cambridge, MA: Belknap Press of Harvard University Press.
- Dewey, J. (1922). *Human nature and conduct; An introduction to social psychology*. New York: Holt.
- DiMaggio, P. J., & Lincoln, J. R. (1995). The new institutionalism in organizational analysis. *Social Forces*, 73, 1147.
- Elder-Vass, D. (2010). *The causal power of social structures: Emergence, structure and agency*. Cambridge: Cambridge University Press.
- Ellickson, R. C. (1991). *Order without law: How neighbors settle disputes*. Cambridge, MA: Harvard University Press.
- Elster, J. (2007). *Explaining social behavior: More nuts and bolts for the social sciences*. Cambridge/New York: Cambridge University Press.
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs: Prentice-Hall.
- Giddens, A. (1979). *Central problems in social theory : Action, structure and contradiction in social analysis*. Berkeley: University of California Press.
- Goffman, E. (1956). *The presentation of self in everyday life*. Edinburgh: University of Edinburgh, Social Sciences Research Centre.
- Goffman, E. (1967). *Interaction ritual; Essays in face-to-face behavior*. Chicago: Aldine Pub. Co.
- Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3), 481–510.
- Gross, N. (2008). *Richard Rorty : The making of an American philosopher*. Chicago: University of Chicago Press.
- Gross, N. (2009). A pragmatist theory of social mechanisms. *American Sociological Review*, 74(3), 358–379.
- Gross, N. (2010). Charles Tilly and American pragmatism. *The American Sociologist*, 41(4), 337–357.
- Hanif, M. (2011, January 5). How Pakistan responded to Salmaan Taseer’s assassination. *The Guardian*. Section G2, p. 4.
- Hedström, P. (2005). *Dissecting the social: On the principles of analytical sociology*. Cambridge/New York: Cambridge University Press.
- Hedström, P., & Bearman, P. (2009). *The Oxford handbook of analytical sociology*. Oxford: Oxford University Press.
- Hedström, P., & Swedberg, R. (1998). *Social mechanisms: An analytical approach to social theory* (Studies in rationality and social change). Cambridge/New York: Cambridge University Press.
- Joas, H. (1985). *G.H. Mead, a contemporary re-examination of his thought* (Studies in contemporary German social thought). Cambridge, MA: MIT Press.
- Joas, H. (1993). *Pragmatism and social theory*. Chicago: University of Chicago Press.
- Joas, H. (1996). *The creativity of action*. Cambridge: Polity Press.
- Joas, H., & Beckett, J. (2001). Action theory. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 269–285). New York: Kluwer.
- Kilpinen, E. (2009). The habitual conception of action and social theory. *Semiotica*, 2009(173), 99–128.

- Korsgaard, C. M. (2008). *The constitution of agency: Essays on practical reason and moral psychology*. Oxford/New York: Oxford University Press.
- Little, D. (1994). Microfoundations of Marxism. In M. Michael & M. I. Lee (Eds.), *Readings in the philosophy of social science* (pp. 479–496). Cambridge, MA: MIT Press.
- Little, D. (2006). Levels of the social. In T. Stephen & R. Mark (Eds.), *Handbook for philosophy of anthropology and sociology* (pp. 343–371). Amsterdam/New York: Elsevier Publishing.
- Little, D. (2012a). Analytical sociology and the rest of sociology. *Sociologica*, 1. doi:10.2383/36894
- Little, D. (2012b). Explanatory autonomy and Coleman's boat. *Theoria*, 27/2(74), 137–152.
- Mead, G. H., & Morris, C. W. (1962). *Mind, self, and society: From the standpoint of a social behaviorist* (Works of George Herbert Mead, Vol. 1). Chicago: The University of Chicago Press.
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge/New York: Cambridge University Press.
- Steinmetz, G. (2007). *The devil's handwriting: Precoloniality and the German colonial state in Qingdao, Samoa, and Southwest Africa*. Chicago: University of Chicago Press.
- Whorf, B. L. (1956). *Language, thought, and reality; selected writings* (Technology Press books in the social sciences). Cambridge, MA: Technology Press of Massachusetts Institute of Technology.



# Chapter 5

## Three Issues in Social Ontology

Philip Pettit

**Abstract** Social ontology gives an account of what there is in the social world, judged from the viewpoint of presumptively autonomous human beings. Three issues are salient. The individualism issue is whether social laws impose a limit on individual autonomy from above; the atomism issue is whether social interactions serve from below as part of the infrastructure of intentional autonomy; and the singularism issue whether groups can rival individuals, achieving intentional autonomy as corporate agents. The paper argues that individual autonomy is not under challenge from social laws, that the achievement of intentional autonomy does indeed presuppose interaction with others, and that groups of individuals can incorporate as autonomous agents. In other words, it defends individualism but argues against atomism and singularism.

### 5.1 Introduction

The ontology of any domain ought to give an account of what there is in that domain, in particular of what there is that counts as interesting from one or another point of view. What counts as interesting from one viewpoint, of course, may not count as interesting from another. The farmer will give one account of what there is to be found in a field, the botanist another, the painter a third. The farmer will focus on the plants in the field; the botanist on the different vegetative life-forms, weeds as well as plants; the painter on the varieties of texture and color that those plants and weeds display against the background of soil and sky.

---

P. Pettit (✉)

Center for Human Values, Princeton University, Princeton, NJ, USA

School of Philosophy, Australian National University, Canberra, ACT, Australia

e-mail: [ppettit@princeton.edu](mailto:ppettit@princeton.edu)

The domain of social ontology comprises the interactions of individual human beings together with the patterns that constrain those interactions or that emerge from them. The interactions relevant, so I shall assume, are those that involve the intentional attitudes of participants: that is, the attitudes that bulk large in the psychology of persons—whether scientific or common-sense psychology—such as belief and desire, judgment and evaluation, intention and policy, emotion and mood (Mellor 1982). Thus the interactions and associated patterns relevant to social ontology include our interactions as friends and foes, consumers and producers, compatriots and foreigners. But they do not extend to interactions that are explicable in wholly sub-personal terms: for example, interactions of epidemiological contagion, pheromonal stimulation or competition for oxygen.

But if this is the domain of social ontology, what is the viewpoint that informs it, making some questions salient, others not? I think that in the traditional and contemporary literature of the discipline—so far as it has a recognizable profile as a discipline—the viewpoint is shaped by an interest in the significance of our social interactions, and of the groups we form in social interaction, for our status as minded agents, guided by intentional attitudes.

There are three main questions that this interest has stimulated and, using terms in a somewhat stipulative sense, I describe them in turn as the individualism issue, the atomism issue and the singularism issue. In this paper I focus on each of these questions in turn and, drawing on earlier work, gesture at some arguments in favor of the positions I adopt. The presentation is excessively condensed but it may offer a useful overview of the field as a whole. I conclude with a brief discussion of the significance of these issues.

## 5.2 The Individualism Issue

### 5.2.1 *History*

The individualist question, which came into prominence only in the nineteenth century—and has perhaps lost its hold on our contemporary sensibility—is whether the forces associated with social life, in particular the forces that social science is liable to unearth, entail that the intentional attitudes posited in personal psychology are not always the forces that move us to action. On at least some fronts we are pawns of unrecognized social forces, so anti-individualists suggest, not the intentionally guided or autonomous agents we take ourselves to be.

Ian Hacking (1991) argues that as social science began to make an appearance in the nineteenth century, it was shaped in great part—and perhaps even called into existence—by the plethora of social statistics that began to appear as a result of the rise of the administrative, bureaucratic state. From about 1820 on the state in various European and other countries began to record and publish figures on, for example, the aggregate incidence of crime, insanity and suicide, poverty, illness

and mortality. And they thereby revealed the rates at which these statistics changed across time and place, if indeed they did change, as well as their correlation or lack of correlation with one another. Many assumed that such rates and correlations would vary more or less at random, given the presumptively random way in which individuals resolve intentional issues and exercise free will. But the data gave the lie to that assumption, revealing unexpected constancies and unexpected degrees of predictability in people's social behavior.

According to Hacking's narrative, the discovery of these unexpected constancies led a great variety of European thinkers to the conclusion that there was a hidden hand at work in social life. This was not the invisible hand of the market that Adam Smith (1976) had charted in the eighteenth century; that is, not a mechanism whereby individual interactions, psychologically intelligible in themselves, would reliably give rise to certain aggregate patterns. And it was not the iron hand of the state: that is, not a mechanism of control intentionally exercised from above. The hand that these thinkers saw at work in the statistical constancies they espied was a much less obvious, and ultimately a much more ominous, force. It was a source of pre-determination in people's behavior that put in question the minded status, and the intentional autonomy, posited in our ordinary psychology and experience of ourselves.

The imagined source of pre-determination was sometimes compared to the silent force of gravity that shapes the movements of astronomical bodies. The idea was that just as the heavenly bodies are forced to move in the patterns that gravity dictates, without any evidence of active push or pull, so we ordinary human beings may be subject to equally silent and equally inescapable forces, being driven unwittingly to display certain socially ordained patterns of behavior. This sort of social determinism was endorsed in a variety of forums. It shaped T.H. Buckle's *History of Civilization in England*, published to great acclaim in 1857. And it assumed a vivid, theatrical form in the vision presented in Tolstoy's *War and Peace*, written between 1863 and 1869. Tolstoy (1972, 1313) writes: 'Ever since the first person said and proved that the number of births or crimes is subject to mathematical laws, that certain geographical and politico-economical laws determine this or that form of government, that certain relations of population to the soil lead to migrations of people—from that moment the foundations on which history was built were destroyed in their essence'. It became impossible, so he suggested, 'to continue studying historical events, merely as the arbitrary product of the free will of individual men'.

This sort of social determinism sponsored the appearance in late nineteenth-century France of a science of society—a sociology, in the name given it by the philosopher, Auguste Comte—that would reveal the laws governing social life. The great protagonist of this movement was Emile Durkheim, one of the founders of sociology as it we know it today. While developing a body of work that is valuable by almost any lights (Lukes 1973), Durkheim nurtured the aspiration to replace the sense of ourselves present in commonsense psychology—and in many scientific versions of the discipline—by displaying the social forces at work amongst us. He took those forces to operate on us coercively, in a way that bypasses our sense of

what we do and why we do it, via a variety of what he called social facts. These include features of our society like the density of population, the norms and rules institutionalized there, the currents of opinion that prevail at any time, and the enthusiasms that occasionally sweep across a group. ‘A social fact’, he says in an account of sociological methodology, ‘is to be recognized by the power of external coercion which it exercises or is capable of exercising over individuals’ (Durkheim 1938, 10).

Durkheim’s 1897 study of suicide—a classic of sociology—illustrates nicely the sort of determinism in which he believed. The statistics on suicide may be wholly unpredictable on the basis of physical or biological or indeed psychological facts, he thinks. But they display a constancy in their relations to ‘states of the social environment’. ‘Here at last’, he says, ‘we are face to face with real laws’ (Durkheim 1951, 299). Asserting the relentless operation of these laws across different cultures and institutions, he comments in conscious irony that ‘Each society is predisposed to contribute a definite quota of voluntary deaths’ (Durkheim 1951, 51). The irony in the use of ‘voluntary’ is underlined by an explicit recognition that his approach is bound to scandalize ‘the zealous partisans of absolute individualism’. ‘For those who profess the complete autonomy of the individual’, he says, ‘man’s dignity is diminished whenever he is made to feel that he is not completely self-determinant’ (Durkheim 1938, 4).

### 5.2.2 *The Issue*

How likely is it that the laws which social science has discovered, or is liable to discover, might give the lie to our sense of ourselves as autonomous agents? Might they suggest that it is a mistake to think we are more or less successfully interpretable in the common psychological terms that we use to make sense of ourselves?

In order to make intentional or psychological sense, we must generally hold attitudes of belief and desire and the like that are rationally intelligible in light of the evidence at our disposal and we must generally act in a manner that is rationally intelligible in light of those attitudes. But we need not be unfailingly rational in these ways. It is part of our psychological understanding that there are various factors, some perhaps yet to be discovered, that cause us to be temporarily irrational, such as when we are preoccupied or upset, or subject to inertia or *idees fixes*. And neither need we see very deeply into the conditions that give rise to psychology-shifting effects as when we fall in love or are shocked by a traumatic experience. In order to be intentionally interpretable to ourselves and one another—in order to count as conversable agents (Pettit and Smith 1996; Pettit 2001)—we need only preserve a general conformity to rational expectations and a capacity, at least in the case of certain failures, to recognize and correct them.

According to anti-individualism, some of the laws of social science—some of the laws actually discovered or liable to be discovered—are downright inconsistent with the intentional or conversable image that we hold of ourselves. They require

people to behave in certain ways and, in particular, to behave in ways that are intentionally unintelligible. The idea is that when social laws require people to respond in a certain manner, then regardless of whether this would make any psychological sense—regardless of whether it would cohere with our belief in their intentional intelligibility—people must respond in that manner. On some interpretations, including Durkheim's own, the laws envisaged may have to be satisfied if the society is to survive and flourish: they are socio-functional necessities. And so the idea is that people are liable to be pushed by socio-functional requirements into performing in a manner that makes little or no psychological sense. They go on the blink as they put themselves, zombie-like, at the service of such necessities.

### 5.2.3 *For Individualism*

There are a number of considerations that argue against anti-individualism, as it has been characterized here. A first is that if we assume that the intentional laws assumed in predication of rationality would hold true in the absence of the social laws envisaged by Durkheim and others—and nothing he says suggests that they wouldn't—then we have to think of the social laws as issuing from a novel sort of force. Vitalists argued that over and beyond the chemical laws governing living things there is a *vis vitalis*—a vital force—that explains why some chemically constituted organisms satisfy biological laws that are chemically unintelligible. And in the same way anti-individualists would have to argue that over and beyond the intentional laws governing agents like you and me there is a *vis socialis*—a social force—that explains why we psychologically organized agents satisfy social laws that are psychologically unintelligible. But just as parsimony argues against vitalism in biology, so it argues against anti-individualism in sociology. We ought to be driven to become anti-individualists only in the presence of undeniable data that cannot be explained in individualist terms.

A second consideration against anti-individualism is that there are no such data available. Even candidate laws that have a Durkheimian cast, and that make a claim to advance our understanding, can be fitted easily within an individualist picture. Assume for argument's sake that it is a social law, for example, that an increase in unemployment gives rise to an increase in crime. If true, this law would tell you something important about the social world. No matter how fully you understood the psychology of individual participants in the society you might not have noticed the regularity it underlines. But the law would not reveal that there is a super-intentional force at work. All that it need posit is that a rise in unemployment, no matter how it is psychologically realized—no matter who lose their jobs and no matter how they feel and think—is likely to give rise in a psychologically intelligible way to an increase in crime.

Whatever individuals become unemployed, and whatever their psychological profile happens to be, the increase in unemployment means that there are more

people than previously with a novel motive to commit crime—to make up for the loss of income—and with a novel opportunity to do so: the increased leisure that unemployment ensures. The increase in unemployment programs for the increase in crime, as we might say, since it means that things are psychologically organized so that, under plausible psychological assumptions about the interaction of motive and opportunity, an increase in crime becomes likely. The increase in unemployment does not produce the increase in crime in a manner that engages a novel sort of force and bypasses people's intentional make-up (Jackson and Pettit 1992a, b; Pettit 1993).

A third consideration against anti-individualism is that not only are its claims inherently implausible and explanatorily unnecessary, they would also run into conflict with psychological tenets that lie at the very center of our web of belief and that it is very hard to imagine giving up. I have in mind the assumptions about our more or less rational character that we mobilize in interpersonal interaction, as we assume that in general we are each conversable—each capable, at least in the normal run, of being reached in conversation. This assumption shows up in our practice of talking to one another about what we ought to believe and desire and do, only despairing of this exercise with the rare individual whom we take to be out of their mind. It is particularly salient in our disposition, absent recognizable excuses, to feel resentment or indignation towards people who fail to register or respond to salient, other-regarding considerations and consequently do harm to us or to third parties (Strawson 1962). It is hard to imagine how we could continue the patterns of exchange and conversation essential to community—and maintain the parallel patterns of self-reflection and self-interrogation in which thought consists—if we gave up on the intentional, conversable image of members of our kind.

### 5.2.4 *Qualifications*

The account given of anti-individualism is motivated both by the history of the approach and by the fact that on this account, anti-individualism has important implications for our status as minded creatures. But I should add that there are many other doctrines that might reasonably claim to be anti-individualist and that no considerations rehearsed here are meant to challenge them.

One is the claim that social science can expand our psychological understanding of ourselves, revealing factors that perturb our normal functioning: this fits comfortably with the commonsense recognition that there is an open variety of emotional and cognitive blocks to optimal performance. Another is the claim that the social entities that come into existence as a result of individual interactions can themselves figure in people's awareness, reciprocally influencing what they do; the appearance of money, for example, can elicit novel sorts of attitude and generate novel sorts of activity. And yet another is the common, if not altogether persuasive claim that for any grand developments associated with particular individuals—say,

the Napoleonic reforms in early nineteenth-century Europe—those developments would have materialized, even in the absence of the individuals involved.

There is clearly no reason in principle why individualists in the sense relevant to our discussion might not endorse such doctrines. But there is another doctrine that may seem to challenge individualist assumptions more directly. According to this theory, there are social laws that are not psychologically intelligible, even if there are none that are psychologically unintelligible. The claim is that certain social laws cannot be derived from psychological laws—strictly, from psychological laws as they operate under various circumstances—not that they require various psychological laws to be false. They transcend psychological laws but do not confound them.<sup>1</sup>

Strictly speaking, this doctrine need not be a challenge to the central individualist claim that social laws do not threaten to compromise our intentional or psychological sense of ourselves. But in any case it is hard to identify a persuasive social law that would resist psychological derivation in the sense required by the theory.

Suppose, by way of constructing such an example, that at an early stage in our evolution whole groups survived or perished in group-group competition; that the groups that survived were ones in which members were disposed under external threat to put aside internal divisions and fight as one against enemies; and that consequently we current human beings have almost all inherited this sort of disposition. Assume that as a result of such a group-selectional history, it is a social law that the members of a society unite against external threat. Might it be plausible to claim that that law is not derivable from psychological laws, as they operate in this or that circumstance? Might it be plausible to hold that this is so, because the law depends on the presence of a disposition that, by hypothesis, is not psychologically intelligible?

While I have no principled objection to the possibility illustrated, and while it does not really threaten the individualist position I hold, I think that the answer to this question must be, no. Any disposition that we inherited from our evolutionary history in the manner illustrated is almost bound to have been registered within our psychological sense of ourselves as a fully intelligible trait. Our intentional psychology has been formed in light of our experience of ourselves and it is surely likely that any evolutionarily established disposition to form certain attitudes under one or another circumstance would have been long identified as typical of our species. This is obviously true, as it happens, with the disposition cited in the

---

<sup>1</sup> In Chapter 3 of *The Common Mind* I describe this sort of doctrine as making the claim that social laws outflank intentional laws rather than overriding them in the manner envisaged by anti-individualists proper (Pettit 1993). The core difference between the overriding and the outflanking doctrines is that whereas adherents of the first take social laws to be inconsistent with psychological laws, adherents of the second allow that they are consistent. Both groups hold that certain social laws fail to supervene on the operation of psychological laws under various circumstances but they make that claim on very different grounds.

example. No one is likely to think that a disposition to make common cause against an external enemy is psychologically underivable and surprising.

This argument is not decisive, of course. It applies only to one putative example of a social law that is not psychologically derivable and intelligible, even if it does not require any psychological laws to be false. But I think that most candidates for the role envisaged are likely to fall to similar considerations. In any case, we need not concern ourselves unduly with the question of whether the theory that such examples would bear out is likely to be sound. For unlike the sort of theory associated with Durkheim and his followers, at least as I have interpreted them, it would not do anything to undermine our status as individually minded agents.

## 5.3 The Atomism Issue

### 5.3.1 History

Where the individualism question is whether people's status as minded, conversable agents survives in the space of aggregate social laws, the atomism issue is whether, on the contrary, that status presupposes a life conducted within the constraints of social relationships. You can have such and such a height or weight quite independently of whether there are any others around but you cannot enjoy prestige or power except in the presence of others. The question here is whether any of the properties associated with intentionality or conversability are more like prestige than they are like height or weight.

Although Aristotle (1996, Bk 1) argued that we human beings are essentially social or political agents, associating this feature with our ability to relate in a deliberative, linguistically mediated way, the atomism issue really came into prominence in philosophical discussion only in the eighteenth century. It became an issue in light of the German Romantic claim, foreshadowed in Rousseau's *Discourse on the Origin of Inequality* (1973), that human beings depend on language for a range of minded capacities and that they depend on society for access to language. This claim became the central theme in the work of objective idealists in the nineteenth century, figuring prominently in the thought of Hegel and his followers.

Thomas Hobbes (1994a) had argued in the early 1640s that language is essential for the appearance of distinctive human capacities, presenting it as the source of what makes human beings special. He formulated this view as an alternative to Descartes's (1985) claim—defended in his *Discourse on Method* of 1637—that language was a sign, not a source, of human distinctiveness; this Descartes took to consist in the presence of thinking substance, *res cogitans*. Hobbes maintained that language is a human invention that changed the nature of its inventors, giving them powers of mind that made them special among animals (Pettit 2008). He argued in particular that without language people would not be able to ratiocinate or reason;



they would be incapable of thinking through theoretical or practical issues. They would not have the capacity exercised paradigmatically by the hunched figure of Rodin's *Le Penseur*.

Hobbes did not think that the language that is essential to performing as a minded human being is essentially social. But from the time when Rousseau began to defend that idea, the atomism issue became a staple of discussion. If language is a construct that emerges only in the interaction of human beings with one another—if, contrary to Hobbes, it could not be the invention of a single person—then human beings are going to depend on interacting with one another for the appearance of the capacities that language underlies, in particular the capacity for thought. It is no longer going to make sense to think that a solitary individual, operating within the space of his or her own consciousness, could achieve the status of a properly minded agent.

The anti-atomists of the nineteenth century rang many changes on this theme. These changes included the Hegelian claim that it is only in interaction with one another that human beings become self-conscious (Hegel 1991). But the changes rung extended more generally to observations on the artificiality of abstracting from social context and treating individuals as the primary units of mind and agency. F.H. Bradley (1876, 173–74), the English idealist thinker, argued in this spirit that ‘the mere individual is a delusion of theory’ and that to ‘know what a man is you must not take him in isolation’.

### 5.3.2 *The Issue*

The question that divides atomists and anti-atomists is whether there are any features essential to human beings—in particular, any feature like the capacity for reason and thought—that depend for coming into existence on the enjoyment of social relations (Taylor 1985). But in order to understand the question properly there are two construals that we should put aside, one of them causal, the other logical.

On the causal construal, the question is whether we human beings depend causally on interaction with others—for example, on interaction with parents and other adults—for the appearance of distinctive mental capacities. Since it would be crazy to deny that we do, this reading of the issue has little or no appeal; it would make atomism utterly implausible and give anti-atomism too easy a victory.

On the logical construal, the question is whether we human beings depend as a matter of logical necessity on interaction with others for the appearance of these capacities. But this reading is equally unappealing, since it would make anti-atomism wholly implausible and give an easy victory to atomism. How might anyone argue that it is inconceivable that creatures like us could enjoy the full range of mental capacities in isolation from one another? To defend such an inconceivability claim would be to maintain that Descartes's image of minded, potentially isolated subjects is not only mistaken, for example, but logically

mistaken: there is no possible world in which people conform to his model. Few if any have ever thought that this was plausible.

I favor a reading of the atomism question that avoids both of these extremes, casting the issue as one of whether we human beings depend in a contingent but non-causal manner on our interacting with one another, or on our ever having interacted with one another, for the possession of distinctive mental capacities. The mode of dependence I have in mind is contingent rather than logical in character and constitutive rather than causal.

Consider your dependence on the presence of suitable antibodies in your blood for the enjoyment of immunity against a certain disease. The antibodies that make you immune do not cause that immunity, as they might cause a distinct, temporally downstream effect; they serve rather to constitute it. Thus you do not have to wait on the antibodies to have a causal effect in order to become immune: you are immune from the moment they are present. And yet the antibodies that make you immune are not logically connected with your immunity. It is possible in principle that you might enjoy immunity by any of a variety of other biological or indeed miraculous means. They constitute your immunity but do so as a contingent matter, not as a logical necessity.

On the construal I favor, the atomism question is whether there is a form of social interaction on which, in a similar manner, we contingently but constitutively depend for the possession of some central feature of human mindedness. If there is such dependence, then the exercise of that capacity will be inherently social in character. And in that sense the anti-atomist claim will have been established.

### 5.3.3 *For Anti-atomism*

Arguments against atomism have to start by picking out a feature of our minded make-up, then, and offer reasons why the presence of that feature presupposes social interaction in a constitutive role. I will sketch an argument that focuses on the capacity to reason and, more basically, to follow a rule. This is not the only sort of argument that might be put forward in support of non-atomism but it has the merit of focusing on a feature of human mindedness that is clearly important to our functioning and that appears to mark us off from other animals. While other animals can reasonably be attributed intentional states like belief and desire and the like, they give little or no evidence of the reasoning or thinking that we human beings conduct, whether on our own or in deliberation with others.

Reasoning in the intended sense may consist in determining on the basis of a rule like *modus ponens* whether a certain conclusion follows from premises already believed. But it may also consist in something much simpler such as wondering whether something not confronted previously is deserving of a familiar name: whether it counts under the appropriate rule of classification as an instance of this or that property or kind. I will concentrate on this latter sort of case, asking whether the rule-following involved in such a basic exercise of classification presupposes

social interaction. I argue that it does, drawing on a response to the problem of rule-following raised by Ludwig Wittgenstein (1958, 1978), particularly in the interpretation of that problem offered by Saul Kripke (1985).

Suppose you grasp the meaning of a term—say, the property ascribed by a certain predicate—and aspire or intend to use the term in an appropriate way: that is, in a way that faithfully tracks the corresponding property. Assuming that you do not grasp the meaning of the term on the basis of definition in other terms—assuming that in that sense the term is semantically basic—you presumably identify the property you mean to track, and so the rule you expect to be guided by, on the basis of examples. Let the term in question be ‘regular’, as that is used of shapes. You will be introduced to the rule you mean to follow by various examples of regular shapes, where these are presented in a suitable contrastive context. Thus the examples used to cue you might be geometrical squares, circles, ellipses, triangles, and horseshoes, where these vary in color, size, font and the like, and are set in contrast to a variety of irregular shapes.

The main problem with rule-following is to explain what fixes the identity of the property or rule that such a finite set of examples is meant to present; in particular, what fixes the identity of the rule in a way that gives you access to that identity: after all, you must know which rule is in question if you are to be able to track it intentionally. There is no doubt, we may assume, that confronted with such a set of examples, and assisted by appropriate contextual priming, you are likely to catch on to the intended pattern. You will form a disposition to extrapolate to other cases, classifying candidate shapes as of a kind or not of a kind with the examples: that is, as regular or irregular. But how could just the formation of such a disposition amount to following a rule? How could it enable you to identify a rule with an indefinitely large extension; to aspire to follow that rule in using the term ‘regular’ across novel cases; and to do so, as rule-following requires, in a way that allows you to recognize that you may get that rule wrong?

The account of rule-following that I favor builds on the assumption that you are indeed likely to form a spontaneous extrapolative disposition in response to a set of examples like those mentioned. But it adds two important elements to that story, one proleptic or anticipatory, the other interpersonal or social. And it claims thereby to be able to explain how the disposition can allow you to identify a rule as something you can aspire to comply with, yet aspire without any absolute guarantee of success (Pettit 1993, 2002).

The proleptic part of the story is that the disposition elicited by the examples enables you, consciously or unconsciously, to conceive of the rule as something you can target as an object of compliance. You can think of it in anticipation as *that* rule, the one that you rely on your disposition to reveal in a case-by-case way.

Imagine, to take a parallel, that you know how to get between two places in virtue of, first, knowing where to go initially as you set out from one or the other end; second, knowing that when you get to that initial landmark you will know where to go next; third, knowing that when you get to the next landmark you will know where to go then; and so on. In such a case you will know *that* route between the two places, the one that is encoded in your disposition to move between

landmarks; and of course you may know that route without being able to draw or describe it.

The idea in the first, proleptic part of the story about rule-following is that in the same way you can know the rule presented in a finite set of examples just by being able to rely on the disposition that is elicited by the examples. You can recruit the disposition to the role of identifying that rule in a case-by-case way and you can use the examples, then, to make that rule available as an object of attention: to pick it out as *that* rule, the one that is going to be salient to anyone with the required disposition.

The anticipatory or proleptic story will not suffice on its own, however, to explain how you can get to identify and follow a rule like that associated with the property of regularity. For all that the story involves in its first stage, you would have no reason to think that you could misidentify the rule on any occasion; you would have no obvious ground for conceiving of your rule-following as fallible. The second, interpersonal part of the story is meant to repair this defect.

The claim in the second part is that in using your extrapolative disposition to identify the rule you mean to follow, you assume that there is something to follow that is available to others too, answering to their dispositions as well as to yours. This means that faced in any instance with a discrepancy between your response and those of others you will naturally balk and look for an explanation that enables you—ideally, enables all of you involved in the divergence—to continue to think that there is something objective you are each meaning to track. The best explanation that is consistent with the objectivity assumption would identify some warping obstacle or oversight on the part of one or another party, so that the rule you aspire in common to follow can be cast as *that* rule, the one that shows up in each of your dispositions when the disposition operates in the absence of such perturbing factors: that is, in the absence of factors that would save the assumption of objectivity and yet explain the divergence.

If anything like this story is on the right lines, then rule-following consists at base in triangulating with others on a presumptively objective pattern, relying on that pattern to be available in virtue of the interplay between individually extrapolative and mutually corrective dispositions. Consistently with the story, it may often be the case that you intentionally and successfully follow a rule in isolation from others. All that is required is that you have had some experience of triangulation in the past and that you acknowledge the relevance of triangulation, if it is available, in the resolution of certain discrepancies. But that is still enough to establish the social character of rule-following. Drop the authorization of others in the identification of basic rules and you will lose any ground for presuming that the rules you mean to track are genuinely objective patterns—patterns that it is possible for you to misidentify.

The story sketched here might be replaced by a story in which each of us means to track the rule identified by our personal, extrapolative disposition, as that operates under presumptively reliable conditions (Blackburn 1984). But the substitute story faces the problem of explaining how we could individually identify such conditions. And even if it were to avoid that problem, it fails to explain how we

can be licensed in assuming, as we routinely do assume, that the pattern we track in the use of a simple term like ‘regular’ is the pattern that others track too. In any event the social story answers much better to our common sense of what transpires in learning the meaning of the terms we use from others in our linguistic community. While there may be a possible world in which human beings each rely on their private, idiolectal resources to identify the basic rules they follow in reasoning, there is little or no ground for thinking that that world is the actual one.

This brisk presentation directs us to one line of argument that makes a good case for anti-atomism. The activity of rule-following and reasoning rests, it would seem, on the availability of a practice of using one another to give ourselves suitable targets of thought: suitable patterns to be guided by in working out what the world we chart in common requires us to say and think in this or that instance. We do not causally depend on the history and availability of such triangulation as we might depend on something distinct from reasoning itself; the dependence is constitutive in character. Nor do we depend on it as a matter of logical necessity; as just noted, there is nothing incoherent in the idea that we might identify and track rules on a private basis. But the sort of dependence involved is still enough to ensure that our capacity to reason and follow rules has a social character. As a contingent but constitutive matter, the ability to reason and follow rules presupposes interaction with others; it is not something that we could enjoy out of society.<sup>2</sup>

There are serious issues raised by the adoption of such a theory of rule-following. For one thing, it means that any basic rule that we follow in reasoning—say, any property we ascribe in the use of a given predicate—will really be an equivalence class of rules that happen to coincide across instances that are in principle accessible to human negotiation. But this is not the place to explore such implications and consider their significance.<sup>3</sup> Let it suffice for the moment that we have found one plausible argument in support of anti-atomism.<sup>4</sup>

---

<sup>2</sup> Suppose that everything in my experience was consistent with having interacted, and being in a position to interact, with others in triangulating on rules. Could I be said to follow rules, even if there were no others with whom I interacted: even if I were a brain in a suitably equipped vat? I do not think that I could be said to follow rules involving properties and objects in a distal world that I share with others, although it might seem to me that I was doing so; after all, there is no such world available to me. At best I might be said to follow rules on a private basis in the proximate world of my neural stimulations.

<sup>3</sup> I consider them in the appendix to the 1996, paperback edition of *The Common Mind*.

<sup>4</sup> Another argument that I might have given starts from the assumption that human beings have a distinctive capacity to use words in speaking for themselves as authoritative spokespersons. Thus I can give an account of certain attitudes or action-plans—perhaps to myself, perhaps to others—treating that account as something more than a fallible report on a par with the report that another might give of me; I can treat it as authoritative in the sense of foreclosing the possibility, should I fail to act accordingly, of excusing myself on the grounds of having misread the evidence about my state of mind. It is plausible that such a capacity to invest my words with authority presupposes the presence of other people and the practice of tying myself to the avowals of attitude and the promises of action that they elicit. Might I have learned to do this by a practice of making avowals and promises to myself? Hardly, since in Thomas Hobbes’s (1994b, Ch 26) words: ‘he that can bind can release; and therefore he that is bound to himself only is not bound’.

## 5.4 The Singularism Issue

### 5.4.1 History

The singularism issue, as I understand it, is whether there are only singular human agents or whether certain groups can also perform in an agential role. We speak loosely of many groups as holding by certain attitudes and performing certain actions. But that need not give the lie to singularism. The issue is whether there are any such groups that constitute agents proper or agents in their own right, as it is often said. In presenting my views on this issue I follow earlier work, in particular work done in collaboration with Christian List (Pettit 2001, 2003; List and Pettit 2002, 2011, 2012).

The singularism issue has a long history, going back to a medieval debate that had been prompted, according to many accounts, by a decree of Pope Innocent IV in 1246 (Kantorowicz 1997). Arguing that a group agent could not be excommunicated—it did not have a soul and could not be sent to hell—Innocent described it as a *persona ficta*. Philosophers and theologians generally took this to mean that such a group was a fictional person or agent, not a person or agent in any real sense (Eschmann 1946). But lawyers developed the view that institutions like universities are examples, not of fictional persons, but of artificial persons: bodies that can act as natural persons act, at least within the context of law, and that count therefore as persons proper (Woolf 1913; Canning 1980). Thus they hailed guilds and towns, parishes and monastic orders, even the Church itself, as artificial persons that could enter contracts, own property, sue and be sued in the courts, and generally bear rights and obligations in the manner of their natural counterparts.

The concept of the artificial person survived in legal usage down to the nineteenth century but at the end of that century it received a great boost from the work of the German legal historian, Otto Gierke, who sought to resurrect the medieval category. He had an enormous influence on English and American legal and political theorists, many of whom took up the case for treating society as an arena of interaction, not just for individual agents, but also for the corporate bodies that they constitute (Hager 1989; Runciman 1997). Those bodies were taken to include the state at the highest level of aggregation but also the guilds and unions, the clubs and associations, the churches and colleges, that individuals constitute in more intimate forms of collaboration. A commitment to the reality of such agents, and to their status as agents proper, was characteristic of a variety of political movements in the early part of the twentieth century—for example, in guild socialism—but people generally retreated from this commitment about the time of World War II, perhaps as a result of an unwanted association with Fascist, so-called corporatist thought.<sup>5</sup>

---

<sup>5</sup> One factor in the demise of this movement is that many of its adherents were given to extravagant statements of its implications, as in Sir Ernest Barker's (1950, 61) talk of 'the pulsation of a common purpose which surges, as it were, from above, into the mind and behaviour of members of any true group'.

### 5.4.2 *The Issue*

Most groups are collections of people united only by a common characteristic or history or location and do not present in any way as agents. They are as varied as the collection of those who are red-haired or over two meters tall, those who come of a certain ethnicity or hold by certain political or religious beliefs, those who live in a particular area or were born at a particular time. But other groups certainly do present as agents, having members who actively join or acquiesce in the collective pursuit of certain goals, for example, and in the collective selection of the means that those who act in the name of group should follow in promoting the goals. The political party that organizes itself to campaign for election, the corporation that sets out to maximize the returns to its shareholders, the church that arranges for the satisfaction of certain proselytizing goals: all such bodies put themselves forward as entities that simulate the performance of individual agents. As individuals embrace a variety of purposes, deliberate about their relative importance and seek to identify the best means for their promotion, so the same is true of the corporate bodies illustrated.

The singularism issue is whether the bodies that simulate individual agency in this way count as agents proper, agents in their own right—whether, in older terminology, they should be treated as artificial persons. There is no agreed criterion of when a corporate body might simulate agency, yet fail to be an agent proper, but I shall take the relevant yardstick to have the following, quite demanding character. A corporate agent will not be an agent proper just insofar as the attitudes it embraces—and so its associated actions—are determined, issue by issue, on the basis of the attitudes of some or all of its members; they are mechanically responsive to corresponding member attitudes. An agent that was responsive in a mechanical, issue-by-issue way to the attitudes of its members would be like an avatar of those members, not an independent agent. Its thinking that such and such or its deciding that so and so would amount to nothing more or less than its members—equally or unequally empowered—having the profile that mechanically generates such aggregate dispositions.

The member responsiveness that would deprive a corporate body of the claim to agency proper can take a variety of forms. Any corporate body will have to form attitudes on the purposes it is to pursue, the priorities that should obtain amongst those purposes, the opportunities available for pursuing them, the best means for doing so in an individual case, and the like. A responsive, and so not properly agential body might fix its attitudes on such issues by majoritarian or non-majoritarian voting among the membership as a whole; by a majoritarian or non-majoritarian process of voting on different issues by different, delegated sub-groups; by one process of voting in the case of one delegated sub-group, another in the case of another; and so on. I am prepared to say that even such a complicatedly responsive group agent has no more claim to be an agent proper than

the group that is controlled by a single dictator and constitutes just a front for that person's purposes and opinions.

### 5.4.3 *For Anti-singularism*

My argument for anti-singularism is that any agent that is organized to simulate agency in the manner of a corporate body must be organized in a manner that rules out mechanical responsiveness and in a way, therefore, that gives it a title to be regarded as an agent proper. There are two claims essential to the argument: first, that any body that simulates agency must be robustly sensitive to the demands of rationality; and second, that the satisfaction of such rationality requirements rules out the satisfaction of the responsiveness requirements. Together those claims establish the conclusion that well-functioning corporate agencies cannot be mechanically responsive to their members and must count as agents proper, agents in their own right.

To be an agent is to have the capacity to endorse goals, to form representations of the environment in response to incoming evidence, and to act according to those representations in pursuit of the goals. To have such a capacity is to form attitudes rationally on the basis of evidence, as we say, to act rationally on the basis of those attitudes, and to maintain only attitudes that are rationally co-tenable. Or at the least it is to be sensitive to any failures in such rationality and to be disposed to put them right.

Taking up the first claim in my argument, then, a group will be able to simulate agency successfully—to mimic the performance of an individual agent—only to the extent that it can satisfy such constraints of rationality or, at the least, be suitably sensitive to failures. And not only must it happen to satisfy those constraints as things actually are; it must also do so robustly. It must be so constituted that as we imagine it being faced with novel evidence on one or another issue, or becoming disposed to embrace a novel goal, we have grounds for expecting that it will adjust so as to maintain a rational, effectively agential profile. Did a group not have this profile then it would not be equipped to act for its purposes reliably: it would often find itself disposed to act in inconsistent ways. And equally it would not be an entity with which we could do business, as in projecting the responses it will make to various overtures, negotiating with it on that basis, agreeing to enter contracts with it, and so on.

The second claim in my argument is that if a group organizes itself to be rationally compliant and sensitive in this robust fashion, and if it confronts an interconnected set of issues on which it has to judge—as any real-world group certainly will—then it cannot organize itself in a mechanically responsive manner. This claim rests on a set of results in social choice theory—specifically, in the branch known as judgment-aggregation theory—that have begun to appear over the



**Table 5.1** A discursive dilemma

	p?	q?	r?	p&q&r?
A judges that	not p	q	r	not p&q&r
B judges that	p	not q	r	not p&q&r
C judges that	p	q	not r	not p&q&r
A-B-C judge that	p	q	r	not p&q&r

last decade (List and Pettit 2002; List and Polak 2010). But it can be illustrated by what I have described elsewhere as the discursive dilemma (Pettit 2001), building on the work of some legal theorists on a related question in law (Kornhauser and Sager 1993).

Suppose that a group of three people, A, B and C, have to make up their views as a corporate agent on four issues: whether p, whether q, whether r and whether p&q&r. And imagine that the group is member-responsive in a majoritarian way, being disposed on any issue to form the judgment supported by a majority of members. The matrix in Table 5.1 shows that majority voting may lead them to judge as a group that p, that q, that r and—on the basis of a unanimous vote—that non-p&q&r. Thus it shows that if the group is to satisfy rational sensitivity, as the simulation of agency requires, then it must breach majoritarian responsiveness.

This example shows that majoritarian responsiveness is not consistent with the rational sensitivity that group agency requires. In order to operate properly as an agent, the members of the group have agree that whenever a majority vote generates a position inconsistent with positions already adopted, as in this case, they should go to a second round of consideration in which, regardless of their individual positions, they decide on which of the inconsistent attitudes to drop.<sup>6</sup> They have to monitor the positions generated over time by the group, taking each vote initially as a straw vote, and act to ensure that in the attitudes finally endorsed the group satisfies the basic requirements of rationality. In short, they have to construct the mind of the group, independently of the minds of its members, so that it is suited for agency. The members might be led under such a procedure to hold as a group that p, that q, that r and that p&q&r, accepting the fact that on the last issue they as a group have to maintain a view that each of them individually rejects.

Our example shows that a group cannot operate on the basis of majoritarian responsiveness and must adopt something like the straw-vote procedure. The various judgment-aggregation results in the literature generalize the claim illustrated. They support the thesis that no matter which form responsiveness assumes, majoritarian or non-majoritarian, centralized or delegated, it is liable to undermine the possibility of a robust form of rational sensitivity. And those results argue for the claim that if a group is to act like an agent, then it cannot be mechanically

---

<sup>6</sup> Might they just agree to let past judgments logically determine the present judgment in any such case, restoring a sort of mechanical procedure? No, because then the attitudes that the group adopted would depend, absurdly, on the order in which the corresponding questions were addressed.

responsive to its members. The group may not follow the straw-vote procedure; that is only one way in which members can give the group they form a mind and an agency of its own. But whatever procedure is followed, the members of every group agent have to do something parallel. They have to allow the needs of group rationality to trump member responsiveness and to prompt the formation of a corporate body that counts as an agent in its own right.

The upshot is that if a group is to simulate agency, as many groups do, then it has to replicate agency; it has to constitute an agent proper and not just an avatar of its members. While this result is surprising, however, it is in no way mysterious. It is not in virtue of any novel force or spirit that individuals come to constitute an agent in its own right but only in virtue of the way in which they organize their collective affairs, in particular the business of generating shared attitudes. The group they form may count as a different agent from its members but it amounts to nothing more or less than the same collection of individuals.

## 5.5 Conclusion

In opening this paper I said that social ontology is naturally guided by an interest in the significance of social interactions for our status as minded agents, guided by intentional attitudes. The positions for which I have sketched a defense support, first, the individualist claim that for all that social laws imply, people are intentional, conversable agents who are sensitive to the demands of rationality and display the modified autonomy ascribed in common sense; second, the anti-atomist claim that nevertheless people depend constitutively on social interaction for the capacity to reason and follow rules that human mindedness presupposes; and third, the anti-singularist thesis that when people come together to behave like a corporate agent, they have to form a collective mind of their own: they cannot tie the attitudes they endorse and enact as a group to the attitudes they hold as individuals.

These three positions in social ontology have important methodological and indeed normative implications (Pettit 1993, Chs 5 and 6; List and Pettit 2011, Chs 3 and 7). Methodologically, individualism argues for seeking only social laws and explanations that make psychological sense; anti-atomism makes a case for grounding psychological explanation in patterns of conceptualization—perhaps displaying cross-cultural variability—established in common across a society; and anti-singularism shows that if we are to make sense of the behavior of a group agent like a corporation or church or state, then among the explanatory strategies explored, we have to make use of the intentional stance we deploy in interpreting individuals (Tollefsen 2002).

Normatively, the three positions have corresponding implications. Individualism helps to vindicate giving priority to the interests of individuals—presumptively, considered as equals—in assessing social arrangements: no institution can make for good that does not make the lives of individuals go better. Anti-atomism suggests that we should reject the traditional idea that the benefits in terms of which to justify

social and political life, establishing its merits in comparison with an anarchistic condition, should be restricted to benefits that individuals could enjoy equally in the absence and in the presence of social relationships. And anti-singularism argues for ascribing real rights and responsibilities to corporate agents, though only a pattern of rights and responsibilities that, as individualism requires, best serves the interests of individuals.<sup>7</sup>

## References

- Aristotle. (1996). *The politics* (S. Everson, Ed.). Cambridge: Cambridge University Press.
- Barker, E. (1950). Introduction. In E. Barker (Ed.), *Gierke, natural law and the theory of society*. Cambridge: Cambridge University Press.
- Blackburn, S. (1984). The individual strikes back. *Synthese*, 58, 281–301.
- Bradley, F. H. (1876). *Ethical studies*. London: Oxford University Press.
- Canning, J. P. (1980). The corporation in the political thought of the Italian jurists of the thirteenth and fourteenth century. *History of Political Thought*, 1, 9–32.
- Descartes, R. (1985). *The philosophical writings* (Vol. 1, J. Cottingham, R. Stoothoff, & D. Murdoch, Trans.). Cambridge: Cambridge University Press.
- Durkheim, É. (1938). *The rules of sociological method*. New York: The Free Press.
- Durkheim, É. (1951). *Suicide*. New York: The Free Press of Glencoe.
- Eschmann, T. (1946). Studies on the notion of society in St Thomas Aquinas, 1. St Thomas and the decretal of Innocent IV Romana Ecclesia: Ceterum. *Medieval studies*, 8, 1–42.
- Goldfarb, W. (1985). Kripke on Wittgenstein on rules. *Journal of Philosophy*, 82, 471–488.
- Hacking, I. (1991). *The taming of chance*. Cambridge: Cambridge University Press.
- Hager, M. M. (1989). Bodies politic: The progressive history of organizational “real entity” theory. *University of Pittsburgh Law Review*, 50, 575–654.
- Hegel, G. W. F. (1991). *Elements of the philosophy of right* (A. W. Wood, Ed.). Cambridge: Cambridge University Press.
- Hobbes, T. (1994a). *Human nature and De Corpore Politico: The elements of law, natural and politic*. Oxford: Oxford University Press.
- Hobbes, T. (1994b). *Leviathan* (E. Curley, Ed.). Indianapolis: Hackett.
- Jackson, F., & Pettit, P. (1992a). In defence of explanatory ecumenism. *Economics and Philosophy*, 8, 1–21. (Reprinted in *Mind, morality and explanation*, by F. Jackson, P. Pettit, & M. Smith, 2004, Oxford: Oxford University Press, 131–162).
- Jackson, F., & Pettit, P. (1992b). Structural explanation and social theory. In D. Charles & K. Lennon (Eds.), *Reductionism and anti-reductionism*. Oxford: Oxford University Press, 97–131. (Reprinted in *Mind, morality and explanation*, by F. Jackson, P. Pettit, & M. Smith, 2004, Oxford: Oxford University Press).
- Kantorowicz, E. H. (1997). *The king’s two bodies: A study in mediaeval political theology*. Princeton: Princeton University Press.
- Kornhauser, L. A., & Sager, L. G. (1993). The one and the many: Adjudication in collegial courts. *California Law Review*, 81, 1–59.
- List, C., & Pettit, P. (2002). Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, 18, 89–110.

<sup>7</sup>I benefitted greatly from comments received on versions of this paper at conferences in the University of Oslo, 2011, the University of Copenhagen, 2012, and the Jean Nicod Institute, Paris, 2013.

- List, C., & Pettit, P. (2011). *Group agency: The possibility, design and status of corporate agents*. Oxford: Oxford University Press.
- List, C., & Pettit, P. (2012). Symposium on group agency: Replies to Gaus, Cariani, Sylvan, and Briggs. *Episteme*, 9, 293–309.
- List, C., & Polak, B. (2010). Symposium on judgment aggregation. *Journal of Economic Theory*, 145(2).
- Lukes, S. (1973). *Émile Durkheim: His life and work: A historical and critical study*. Harmondsworth: Penguin Books Ltd.
- Mellor, D. H. (1982). The reduction of society. *Philosophy*, 57, 51–75.
- Pettit, P. (1993). *The common mind: An essay on psychology, society and politics* (Paperback ed. 1996). New York: Oxford University Press.
- Pettit, P. (2001). *A theory of freedom: From the psychology to the politics of agency*. Cambridge/New York: Polity/Oxford University Press.
- Pettit, P. (2002). *Rules, reasons, and norms: Selected essays*. Oxford: Oxford University Press.
- Pettit, P. (2003). Groups with minds of their own. In F. Schmitt (Ed.), *Socializing metaphysics*. New York: Rowan and Littlefield, 167–193.
- Pettit, P. (2008). *Made with words: Hobbes on language, mind and politics*. Princeton: Princeton University Press.
- Pettit, P., & Smith, M. (1996). Freedom in belief and desire. *Journal of Philosophy*, 93, 429–449. (Reprinted in *Mind, morality and explanation*, by F. Jackson, P. Pettit, & M. Smith, 2004, Oxford: Oxford University Press).
- Rousseau, J.-J. (1973). *The social contract and discourses*. London: J.M. Dent & Sons Ltd.
- Runciman, D. (1997). *Pluralism and the personality of the state*. Cambridge: Cambridge University Press.
- Smith, A. (1976). *An inquiry into the nature and causes of the wealth of nations*. Oxford: Oxford University Press.
- Strawson, P. (1962). *Freedom and resentment and other essays*. London: Methuen.
- Taylor, C. (1985). *Philosophy and the human sciences*. Cambridge: Cambridge University Press.
- Tollefsen, D. (2002). Organizations as true believers. *Journal of Social Philosophy*, 33, 395–410.
- Tolstoy, L. (1972). *War and peace*. London: Heinemann.
- Wittgenstein, L. (1958). *Philosophical investigations*. Oxford: Blackwell.
- Wittgenstein, L. (1978). *Remarks on the foundations of mathematics*. Oxford: Blackwell.
- Woolf, C. N. S. (1913). *Bartolus of Sassoferrato*. Cambridge: Cambridge University Press.

# Chapter 6

## Collective Responsibility and Group-Control

András Szigeti

**Abstract** Collectives are more or less structured groups of human beings. Responsibility-collectivism is the view that the moral responsibility of at least some such collectives is something over and above the combined moral responsibility of individual group members. This paper focuses on one of the key conditions of responsibility: the requirement of control. It is plausible that this requirement also applies to collective agents and so collective responsibility presupposes group-control. Responsibility-collectivists have often tried to unpack the idea of group-control as non-causal control. I argue that non-causal control is not an admissible basis for attributing responsibility. Only causal group-control is. This is because non-causal group control does not provide the right kind of information regarding the ancestry of a certain outcome. In the second half of the paper, I discuss the difficulties which arise for responsibility-collectivism if one understands group-control as causal group-control. One of these difficulties is whether causal group-control is consistent with ontological individualism. The second concerns the relationship of group-control and individual control. I argue that the first difficulty is manageable, but only at the price of having to accept a solution to the second difficulty which runs counter to the original aim of the responsibility-collectivist of characterizing irreducible collective responsibility as compatible with individual responsibility. Worse still, responsibility-collectivists may have to choose sides in other areas of social ontology as well. This further raises the price of this position.

---

A. Szigeti (✉)  
Department of Philosophy, University of Tromsø, Norway  
e-mail: [andras.szigeti@uit.no](mailto:andras.szigeti@uit.no)

## 6.1 Introduction

Collectives are groups of human beings ranging from formal organizations (e.g., corporations, intergovernmental bodies) to associations of various kinds (e.g., clubs, teams, armies) and random assemblages of people (e.g., victims of an environmental disaster) (Held 1970; French 1979, 1984; May and Hoffman 1991). Responsibility-collectivism is the view that the moral responsibility of at least some such collectives is something over and above the combined moral responsibility of individual group members.<sup>1</sup> I will refer in the following to such supposedly irreducible collective responsibility as *robust* collective responsibility (or sometimes the responsibility of the group *qua* group).

One consideration typically invoked in support of collectivism about responsibility is that without attributing responsibility to collectives *qua* collectives there will be a “deficit in the accounting books” (Pettit 2007a, 194; see also French 1979, 207; Kutz 2000, 113; Copp 2006, 216, etc.). That is to say, nobody will be called to task for many kinds of harms (or praised for the benefits) the source of which appears to be the existence of collectives. This seems worrying in view of the impact corporations, governments and international organizations can have on our lives.

The said deficit can arise, it is argued, because the responsibility of the group *qua* group does not necessarily reduce to the aggregate responsibility of individual members of the group (and also does not translate into the aggregate plus the individual responsibility of *non*-members). In the most dramatic cases, it is possible that there is no responsibility to be allocated at the individual level at all, and yet the group is morally responsible as a group. In general, responsibility-collectivists think that the following can be true: You have allocated all the moral responsibility there is to be allocated at the individual level, and still you have not allocated all the responsibility there is to be allocated. If you want to allocate all the responsibility there is to be allocated, you have to ascribe responsibility to the group as a group as well. Hence the need for the concept of robust collective responsibility.

Strictly speaking the responsibility-collectivist makes two claims. First, that in certain collective contexts “responsibility-voids” (to borrow an expression from Braham and van Hees 2011) can arise. This can happen, among others, due to formal characteristics of collective decision-making processes, or because responsibility-undermining excuses apply to the participating individual agents (Copp 2006), or because the moral significance of individual contributions does not “add up” to the wrongness of the collective outcome, e.g., when the relevant effect is massively overdetermined and/or the individual contribution is very small

---

<sup>1</sup> The version of responsibility-collectivism I will focus on holds that it is the same kind of moral responsibility we attribute to collectives as to individuals. That is, roughly, retrospective desert-based responsibility implying blameworthiness and praiseworthiness and justifying certain normative responses such as punishment and Strawsonian reactive attitudes such as resentment and guilt. Pettit (2007a) and Shockley (2007) explicitly accept this condition.

(Kutz 2000; Lawson 2012). The second claim is that at least some of these voids should be filled by allocating responsibility directly to the collective.<sup>2</sup>

In the present context, ought certainly implies can, so if responsibility-collectivists are to make good on their claim that we should sometimes ascribe responsibility to the group *qua* group, then they have to show that we *can* ascribe responsibility to the group *qua* group. The task, in short, is to show that groups can be proper addressees of responsibility-ascriptions, that some groups at least are “fit to be held responsible” (Pettit 2007a). Since there is a good case to be made that moral responsibility presupposes agency, this task will involve, among others, showing that the collectives can be agents.

But agency is not sufficient for moral responsibility – further conditions have to be met for an agent to qualify as an addressee of (moral) responsibility-ascriptions. In this paper, I will focus on one of the key conditions said to be required for collective responsibility. This is the requirement of control. Clearly, any agent will have to be able to have some control over whatever<sup>3</sup> she is found responsible for. So collective agents too have to be able to exercise some control over whatever it is they are found collectively responsible for.<sup>4</sup>

This is not disputed by parties to the debate (Sect. 6.2). However, for reasons to be explained below (Sect. 6.3), responsibility-collectivists have often tried to unpack the idea of group-control as *non-causal* control (see esp. Pettit 2007a; Shockley 2007). I will argue that non-causal control is not an admissible basis for attributing responsibility (Sect. 6.4). My suggestion is to accept that moral responsibility requires causal control. If we therefore take the idea seriously suggested by the responsibility-collectivist that collectives can be morally responsible agents in the same sense as individual agents, which is the responsibility-collectivist’s avowed aim, then the relevant collectives are to meet the causal condition as well. Non-causal group-control cannot justify ascribing responsibility to the collective. Only causal group-control would.

<sup>2</sup> One feasible strategy available to the responsibility-individualist is to accept the first claim, but reject the second. Perhaps we just have to resign ourselves to the fact that collectively brought about harms can also sometimes be of a kind for which nobody is responsible, comparable to harms inflicted by nature, even though they are in some sense “man-made” (see Sziget 2014).

<sup>3</sup> Responsibility for what, i.e., actions or outcomes of actions? When necessary to specify, I will focus on responsibility for outcomes for simplicity’s sake. Elsewhere I will follow Pettit (2007a) who does not distinguish between responsibility for actions and responsibility for outcomes. The distinction may be relevant to the requirement of control insofar as the necessity of causal control is widely held for outcome-responsibility (actions causing outcomes). By contrast, some libertarians (simple indeterminists) do not accept that agents need to cause the actions for which they are responsible (see Goetz 1988).

<sup>4</sup> Naturally, one can also argue *against* the collectivist position using the dependence of responsibility on agency as one’s point of departure. If no groups of human beings can be agents and moral responsibility presupposes agency, then individualism about responsibility would follow (for such responsibility-individualist arguments, see Miller and Mäkelä 2005; Haji 2006; McKenna 2006). Shockley (2007) denies that collective moral responsibility presupposes collective agency, but accepts the control-requirement.

In the second half of the paper (Sects. 6.5, 6.6, 6.7, and 6.8), I will also discuss the difficulties that arise for responsibility-collectivism if one understood group-control as causal group-control (Sect. 6.5). One of these difficulties is whether causal group-control is consistent with ontological individualism. The second concerns the relationship of group-control and individual control. I argue that the first difficulty is manageable (Sect. 6.6). However, it is only manageable at the price of having to accept a solution to the second difficulty that runs counter to the original aim of the responsibility-collectivist. This aim was to characterize irreducible collective responsibility as compatible with individual responsibility (Sect. 6.7). Worse still, responsibility-collectivists may have to choose sides in other areas of social ontology. This further raises the price of this position (Sect. 6.8).

## 6.2 Responsibility-Collectivism and Group-Control

There is an obvious worry about responsibility-collectivism. Collectivists<sup>5</sup> would not want to deny that individual human beings act when Tim, Dick or Harry signs the contract, fires the gun, votes for awarding tenure, etc., and not IBM, the execution squad, or the committee. Thus Pettit says, for example: “Whatever a group does is done by individual members on behalf of the group and is done intentionally by those individuals” (Pettit 2007a, 191). Formal organizations are typically seen as the strongest candidates among the different types of collectives to qualify as full-blown collective agents (French 1979). Yet everybody seems to agree that even in the case of formal organizations: “whatever the organization does is done by its members in its name” (Ylikoski 2012, 32).

Nevertheless, the collectivist suggests that it is sometimes appropriate to describe some individual actions as actions *controlled* by the group. This would also explain *why* we should hold the collective responsible in the relevant cases. It is plausible to say that the collective is to incur responsibility *qua* collective for whatever it controls *qua* collective (provided the collective also meets other criteria of “responsibility-fitness”).

When saying that an agent is in control of *X*, I mean that it is up to the agent whether *X* will obtain or take place. This is in accordance with how this concept is standardly used in the literature (see O’Connor 1995; Fischer and Ravizza 1998). If an agent is in control of her action, then it is up to her whether she will undertake that particular action. If an agent is in control of an outcome, then it is up to her whether that outcome will take place. We can say that at the very minimum when an

---

<sup>5</sup> In this chapter, the unadorned term “collectivist” or “collectivism” is always to be read as short for responsibility-collectivism or responsibility-collectivist as I defined the position in the first paragraph.



agent is in control the pathway to future states of the world lead through her agency.<sup>6</sup> The world unfolds in the way it does because of her agency.<sup>7</sup>

Collectivists do not wish to deny either that even when the individual action is controlled by the group the individual can act freely and voluntarily. The kind of group-control the collectivist is after does not exclude that the individual who is controlled by the group is capable of resisting group-control. For example, she may refuse to implement the group's policies or decisions and she is individually responsible for that choice. But if this is true, then the individual retains some control over her actions even when those actions are controlled by the group (see, for example, Pettit 1996, 2007a).

It is easy to see why the responsibility-collectivist would want to insist on this point despite the difficulties it gives rise to. Intuitively, many of the paradigmatic cases of robust collective responsibility discussed by the collectivist – e.g., employees acting in the name of their companies or organizations – do not seem to involve diminished individual responsibility (see Kutz 2000). That is, in many of these cases, the individual agents appear to be fully fit to be held responsible themselves, and that is how their individual actions will be assessed morally. In general, the aim of the responsibility-collectivist is not to fill one “gap in the accounting books” by creating another. It is hoped that responsibility-allocation will not turn out to be a zero-sum game as regards the distribution between the individual and collective levels. Rather, the idea is to show that even when all individual responsibility has been allocated, it is possible that there will still be something the collective can be morally responsible for as a collective. Moreover, even in those special cases when there is no individual responsibility to be allocated, responsibility is not allocated to the collective *faute de mieux*, i.e., *only because* no individual is responsible.

It follows that the notion of group-control crucially hinges on the question of how group-control is related to individual control. The collectivist needs to explain how the group can control certain actions of its individual members whereby such group-control (i) does not undermine the individual's control over her actions, but (ii) nevertheless is such that the group also becomes responsible *as a group* for those actions. The collectivist believes that some groups in fact control at least some of the actions of participating individuals in a sense that meets both conditions (i) and (ii). The suggested collectivist solution I will consider here first (Sects. 6.3 and 6.4) purports to show that the group can control actions of its members *non-causally* by arranging that certain individuals perform certain actions and that they perform those actions in certain ways (Pettit 2007a, 191; Shockley 2007). Such “programming” is not causally efficacious itself. Only individual actions are

---

<sup>6</sup> I say “at the very minimum” in order to accommodate the semi-compatibilist argument that control as defined above is all that is required in terms of control for the agent to be morally responsible, whereas regulative control associated with the ability to do otherwise is not (see Fischer and Ravizza 1998). Note that both kinds of control are described by semi-compatibilists as *causal*.

<sup>7</sup> I will define causal control in Sect. 6.5 below.

causally efficacious.<sup>8</sup> Nevertheless, the program is somehow supposed to ensure that the individual actions will in fact produce the effects foreseen by the given collective.<sup>9</sup>

For example, many groups may be ascribed a constitutionally-enshrined or tacitly agreed-upon purpose or policy (e.g., help the poor, conquer the Japanese market, win the Olympics). Other groups may not have long-term policies, but can still be characterized as being committed to a collective judgment on some normative or factual issue. These collective purposes, policies or judgments can be seen as the “program” which the relevant individual actions implement. Emphatically, the influence of the program on the individual actions is not supposed to be causal<sup>10</sup> and should be consistent with individual control. Nevertheless, when some individual does implement the program, then the group’s impact on the individual action can be decisive enough to make the group responsible as a group as well. Or so the collectivist argues.

Two closely related questions arise with regard to this idea. First, in what sense does the “program” *control* the relevant individual actions without causing them? And second, how can the “program” be the basis for attributing responsibility to the group *qua* group if it does not cause the relevant individual actions? I reconstruct the collectivist answer to these questions in the next section before turning to my objections.

### 6.3 Non-causal Group-Control

The idea of non-causal group-control is an application of a general theory about higher-order explanations in the special sciences, folk psychology and elsewhere (Jackson and Pettit 1990). That theory offers a novel, non-reductionist solution to the problem of the causal relevance of higher-order properties in such explanations.<sup>11</sup> The suggested solution is that higher-order properties, while not being

---

<sup>8</sup> More precisely, the physical realizations (which are property-instantiations) of the individual actions. I ignore that complication here.

<sup>9</sup> In addition, the responsibility-collectivist also needs to show that group-control is not just control by some other individual or aggregate of individuals. This task involves, among others, showing that when the group is in control the relevant individual actions implement some autonomous group attitude held by the group *qua* group. I will not discuss the difficulties associated with this issue in this paper (but see Szigeti (2014) and footnote 18 below).

<sup>10</sup> Having said that, it should be noted that, despite their insistence that “programming” is a strictly non-causal process, the language used by advocates of this theory is worryingly causal at times. They talk about the program or arrangement as “ensuring” or “making it probable” (Jackson and Pettit 1990, 114) that basic, causally efficacious factors will bring about the pertaining effect (for the same causal language, see also Jackson and Pettit 1992a, b).

<sup>11</sup> In the following, I will mostly talk about higher-order/lower-order properties, not events. It will be assumed that the problem and possible solutions would be about the same for higher-order/lower-order events as well.

causally efficacious themselves, constitute an indispensable part of causal explanations. These higher-order properties are therefore causally relevant.

Some such theory of higher-order properties is required in order to account for the problem of how to square with one another the following assumptions, all of which appear to be *prima facie* plausible: (i) only lower-level (physical) properties are causally efficacious, (ii) higher-level properties (such as those studied by the special sciences, e.g., psychological properties) are not reducible to lower-level properties, but nevertheless (iii) supervene on physical properties, and (iv) there are significant explanatory generalizations directly linking higher-level properties to one another (e.g., about causal relations between psychological states), and (v): if a lower-level property is causally sufficient for some effect, then typically (that is, barring cases of overdetermination) no property supervening on that lower-level property can be causally efficacious in bringing about that effect (the so-called *exclusion principle*). The question is: if (v) is accepted, then how can we save (iv), i.e., establish the non-redundancy of generalizations in terms of higher-level properties, and hence the autonomy of the special sciences? The “programming-perspective” has been proposed to solve this problem.<sup>12</sup>

The explanatory indispensability of higher-order properties is said to be demonstrated by the fact that if we ignored them we would miss out on essential information regarding the causal process. The causally relevant piece of information is this: if the given causally efficacious factor had not caused the effect, another similar factor would have (Jackson and Pettit 1992b). This counterfactual obtains because of the “programming”. For example, the boiling of water in a closed container may be indispensable to explaining why the wall of the container cracks, even though the causally efficacious factor was in fact the collision of a given H<sub>2</sub>O molecule (or micro cluster of H<sub>2</sub>O molecules) of a certain momentum with the glass. The macro property of boiling is causally inefficacious and supervenes on the basic, causally efficacious micro property. At the same time, that the water is at boiling temperature “programs for” the collision of one or the other molecule (or molecular cluster) with the wall of the container. So the boiling is causally relevant to the explanation of why the container cracks.

This general theory is also applied to group-control (see esp. Pettit 2007a; Shockley 2007; List and Pettit 2011). According to the collectivist, a group can control actions of its members by non-causally “programming” for those actions. We have already provided a preliminary sketch of how this is supposed to happen in the previous section. Groups are committed to certain goals or policies or a collective judgment. These commitments constitute the “program”. In addition, many groups also have a clear division of labour for the implementation of these commitments. As a result, the group will have structured or “programmed for” the actions of the participating individuals by making sure that if one participating individual fails to perform his role within the given division of labour, then some

---

<sup>12</sup> A possible alternative is to deny (i). See Woodward 2008; Menzies 2008; List and Menzies 2010; Shapiro 2012. More on this alternative in Sect. 6.6.

other member will step in. Again, the claim is not merely epistemic: the decisive counterfactual is said to obtain *because of* the “programming”. Additional evidence offered in support of non-causal group-level control is that changing membership need not disrupt the functioning of the group. This is possible because the group will ensure in the “programming” sense that the new members will continue to promote the group functions.<sup>13</sup>

## 6.4 Non-causal Group-Control and Collective Moral Responsibility

I will not discuss the general theory about the causal relevance of higher-order properties here. The point I would like to make is that even if that theory can justify distinguishing among different levels of explanation (Jackson and Pettit 1992a), it does not help in grounding robust collective responsibility.

Since the “program” is itself causally inefficacious, it makes no direct contribution to any given outcome (*pace* Shockley 2007). This is granted by advocates of the “programming account”. The boiling does not break the glass – some molecule does. The corporate does not sign the contract – some manager does. The reason why the “program” is said to be causally relevant is that it provides information about how the causal history leading up to that event *could have gone*. For example, given that the water was boiling, if not this molecule or molecular cluster, then another would have shattered the glass. But this counterfactual information about possible alternative causal histories is irrelevant as regards the moral responsibility of the collective.

Consider the case of punishment and other sanctions based on statistical profiling: Given certain psychological-cum-sociological profiles, it can be highly likely that individuals of that profile will commit certain crimes. We can say that the profile “programs for” the committing of those crimes: if not this particular person of that profile will commit the crime, then the next one with the same profile very likely will. In fact, some argue that given the likelihood that individuals of a distinct profile will commit certain crimes, profiling in itself can constitute a sufficient justification for the application of certain sanctions including punishment. This is of course a highly contentious issue.

However, *even if* we were to accept this argument about profile-based sanctioning, note the following. It makes no sense to say that we should ascribe responsibility *to the profile* on account of its explanatory role, namely that it “programs for”

---

<sup>13</sup> Furthermore, the collectivist argues that there are good reasons why the group *should* exercise such control over its members. Group-control may be required to ensure both the diachronic and synchronic rationality of collective behaviour (List and Pettit 2002, 2011; Pettit 2003, 2007b; List 2006). Of course, collectivists also acknowledge that groups can also fail to perform their functions or can even fall apart completely.

the committing of the particular sort of crime. Crucially, this worry is conceptual, not moral. One cannot hold an abstract statistic morally responsible. That is not a question of moral justifiability, but a matter of avoiding a category mistake.

By the same token, one might perhaps argue that we should impose certain sanctions on members of a collective provided membership in the collective is demonstrably correlated with an increased likelihood of whatever morally or legally objectionable actions. And yet even if this (questionable) argument was to be accepted, it would not follow that we hold the *collective* robustly responsible in such cases. “Programming” is an abstract metric characterizing certain distributions or topologies.<sup>14</sup> Consequently, it cannot be held morally responsible. Again, doing so would simply amount to a category mistake.

It may be replied here that there is a crucial difference between the profile in an example such as the one above and the “program”. The “program” has been “written” by the collective, it may be said. Think of a corporate policy or some binding rule adopted by (say) a consensus or a majority vote.<sup>15</sup> The program then could perhaps be interpreted as a set of instructions or guidelines given by the group to the individual agent who implements the program or some part of it. Collective control of the relevant individual actions would be ensured through these instructions or guidelines.

This objection is misguided, however. First, as clear-headed collectivists themselves acknowledge “[j]ust as anything that the group does is done by members, any instructions that the group gives are given by one or more members” (Pettit 2007a, 189). So the attempted reply by the collectivist would only push the problem back one stage. The problem about collective control would now have to be posed about those intentional actions which amount to the giving of instructions to the implementing individual.<sup>16</sup>

Second, as regards the question of control it is actually irrelevant who put the “program” in place. The question is why the collective should be responsible as a collective when some individual implements that “program”. Now, it may be true that if this individual decided not to implement the “program”, then another individual would have. The problem is that the “programming account” fails to explain what role the collective *qua* collective plays in getting this or that particular individual to implement the program.

---

<sup>14</sup> Note that collectivists themselves write that in the case of the cracking flask, for example, the higher-order property of boiling is an “abstract statistic” (Jackson and Pettit 1992b, 117; the term used in the same context in Jackson and Pettit 1990, 110 is “aggregate statistic”).

<sup>15</sup> Some responsibility-collectivists also emphasize the importance of other, less formal ways of adopting a “program” such as via a shared culture or common norms or goals or feelings of solidarity (May 1991; Shockley 2007).

<sup>16</sup> In general, the collectivist is to avoid circularity. It cannot be argued that what distinguishes a collective profile as in the above example from genuine group-control is that group-control is exercised by the collective *as an agent* (pace Pettit 2007b). Collectivists themselves accept that for a collective to qualify as an agent it already has to possess group-control.

As noted, according to the collectivist “programming” by the collective should be compatible with individual control on the part of those who implement the program (Jackson and Pettit 1992b, 130). However, an immediately striking difference to molecules in the boiling case is that in many cases the individual agent in a collective is free to refuse to carry out the program and voluntarily complies when she does. Being causally inefficacious, the “program” merely identifies what the required tasks are and ensures that each task will be carried out by *some* member of the collective: “maybe these or maybe those” (Pettit 2007a, 192). But if that is the case, then it is hard to see how the collective is supposed to control individual action *qua* collective.<sup>17</sup>

To conclude, given that the “program” itself is causally inefficacious, the “program” merely *predicts* but does not make it the case or “ensure” that some individual will implement the “program”. Consequently, even if it makes sense to talk about there being a “program” in place in some collectives this does not make the collective responsible for any individual actions, not even if other conditions of collective responsibility could be met.<sup>18</sup>

## 6.5 Causal Group-Control?

At this point, the obvious question is: why not save responsibility-collectivism by adopting a causal understanding of group-control? Recall the definition of control above (Sect. 6.2): when an agent is in control of  $X$ , say an outcome, it is up to the agent whether  $X$  will obtain or take place. Causal control then means that the dependence of the outcome on the agent will be causal. The agent will influence what will happen by causally interacting with the world.

It is quite plausible that there is an intimate connection between causal control and moral responsibility. At least when it comes to individual action, there appears to be a near consensus that I am not morally responsible for an outcome unless I had

---

<sup>17</sup> It is worth noting as well that the “program” does not even make a counterfactual contribution in the sense of making sure that a Frankfurtian back-up plan would be executed. It is not the case that should this individual fail to comply, the program would ensure that the program is nevertheless implemented by *making* someone else do it. So if individual agent  $I_1$  fails to perform her task and some other individual agent  $I_2$  steps in and performs the action instead, then that will be once again  $I_2$ 's choice given that program is causally inefficacious.

<sup>18</sup> Including the condition, extensively discussed by responsibility-collectivists (see esp. List and Pettit 2011), that the group has to be able to hold autonomous judgments which can come apart from judgments of individual members. I have argued elsewhere (Szigeti 2014) that it does not follow from the possibility of group judgments being autonomous in this sense that the group *qua* group is responsible for them. This because either some individual or nobody is responsible for these collective judgments. By contrast, my point *here* is that even if group judgments can be autonomous in this sense, this does not mean that the group controls individual actions when individuals implement those judgments.

causal control over that outcome (Fischer and Ravizza 1998; Moore 2009).<sup>19</sup> It is causal control which ensures that I am hooked up to the world so that I at least stand a chance of impacting on what happens out there (Gilbert 2002; Sartorio 2004<sup>20</sup>). Lacking such control, the necessary link is severed. Thus, other things being equal, merely approving of *X* or thinking that it is a good thing that *X* happens will not make one responsible for *X* if approval has no causal effect whatsoever. Allocating causal responsibility is of course a difficult matter because a number of agential contributions may be necessary for a given outcome. The point here is only that if an agent is not in a position to make any causal contribution to bringing about *X*, then that agent will not be morally responsible for *X*. But non-causal control appears to make collective agents causally impotent in this way.

So the question is: Since the requirement of causal control is so plausible, and if my objections are on the right track, “programming” cannot be sufficient for collective moral responsibility even if all other conditions of “responsibility-fit-ness” are met, why not unpack the concept of group-control as causal control? In the following two sections, I want to point to two problems which may be seen as standing in the way of this strategy. The first of these seems manageable. But solving that problem comes at the price of having to accept a response to the second problem which I believe will be found unappealing by many responsibility-collectivists.

## 6.6 Causal Group-Control and Ontological Individualism

The first problem is that if group-control is cashed out as causal control, then this entails attributing causal efficacy to collectives (or causal powers to use a slightly different idiom). That is, groups *qua* groups will have to be treated as causally efficacious factors in causal explanations. The question is whether doing so would saddle the responsibility-collectivist with unwelcome ontological commitments such as ontological holism. Almost everyone today, and certainly responsibility-collectivists discussed in this chapter, embrace ontological individualism (Zahle 2007; Epstein 2009). This is the view that group-level properties supervene on properties of individuals. Fix properties at the individual level and you will have fixed all the group-level properties as well (List and Pettit 2011, 89).

The worry then is that understanding some collectives as causally efficacious factors may not be consistent with ontological individualism understood in this way. This worry is based on the exclusion principle mentioned earlier. It follows from this principle that if properties of collectives are indeed supervenient on

---

<sup>19</sup> As noted earlier, simple indeterminist libertarians question the requirement of causal control for the agent-action as opposed to the action-outcome relationship. However, it seems that they too would accept the requirement of causal control for the action-outcome relationship.

<sup>20</sup> Admittedly, Sartorio denies that moral responsibility entails causal responsibility. I am not sure whether she would also deny the requirement of causal control as defined above.

properties about individuals, then those properties cannot be causally efficacious. The “programming account” has been suggested precisely to avoid both of two seemingly unattractive positions: the Scylla of denying the exclusion principle and the Charybdis of denying the relevance of higher-order properties to causal explanations. As we have seen, the suggested solution consisted in the claim that higher-level properties, i.e., the “program”, can be causally relevant *without* being causally efficacious. If so, then collectives too can be causally relevant without being causally efficacious. We have also seen, however, that denying the causal efficacy of collectives creates a serious problem for responsibility-collectivists because if collectives are not causal efficacious it is hard to see how they can control outcomes.

This problem generalizes.<sup>21</sup> If higher-level properties are not causally efficacious, then they will be causally inert, that is, epiphenomenal. But if higher-level properties are epiphenomenal, then it is hard to see how they can nevertheless be relevant to causation (see Crane 2008, 189). Acknowledging this problem, many bite the bullet and argue that we should attribute causal efficacy to higher-level properties. Such solutions are based on replacing what many see as a “tendentious conception of causation” (Shapiro 2012) underlying the original formulation of the exclusion principle.

It is suggested that if we adopt a counterfactual account of causation, there is no principled obstacle to attributing causal efficacy to higher-level properties (Yablo 1992; Woodward 2003, 2007, 2008; List and Menzies 2009, 2010, etc.).<sup>22</sup> Setting aside a number of complications and refinements discussed since the classic formulation of the account by David Lewis (1973), the gist of the counterfactual account is that for two distinct events<sup>23</sup>  $c$  and  $e$ ,  $e$  causally depends on  $c$  iff: (i) if  $c$  were to occur  $e$  would occur ( $C \square \rightarrow E$ )<sup>24</sup> and (ii) if  $c$  were not to occur  $e$  would not occur ( $\neg C \square \rightarrow \neg E$ ). It will be especially important to keep in mind for what follows that causal dependence is formulated in terms of the truth of *both* the negative and a positive counterfactual (i) and (ii) (see Lewis 1973, 563).<sup>25</sup>

This general solution can be used to attempt to capture the causal role of collectives as well. It is clear that many relevant counterfactuals involving

---

<sup>21</sup> Which is not surprising since, as noted above, the “programming account” is a general proposal for distinguishing the causal relevance of higher-level properties from the causal efficacy of lower-level ones.

<sup>22</sup> Note that libertarian agent causationists have argued that attributing causal efficacy to certain emergent properties does not even require repudiating the conception of causation as generation or production (see esp. O’Connor 1994, 1995).

<sup>23</sup> Note that in terms of this definition the causal relata for Lewis are events, whereas in the rest of the paper I talk about properties, or better property-instantiations, as causal relata. I believe we can ignore this difference for the purposes of this paper.

<sup>24</sup> Whereby  $C$  and  $E$  are propositions referring to the occurrence of the corresponding events or the instantiations of properties.

<sup>25</sup> Whereby the truth of counterfactuals is interpreted, as is standard practice, in terms of a similarity relation between possible worlds.



collectives will come out true. Consider claims such as “had the mob not stormed the Bastille, it would not have fallen” or “if the National Bank raised interest rates, inflation would fall”. What is more, there appears to be a robust counterfactual dependence between properties of collectives and corresponding effects in that both negative and positive counterfactuals of the type (i) and (ii) above appear to hold. If so, then we could regard collectives as causally efficacious using the counterfactual account (see esp. List and Menzies 2009, 2010).

At the same time, it is not disputed by advocates of higher-level causation that causally efficacious higher-level properties remain supervenient on lower-level properties. Thus in the case of collectives too, when the properties of collectives are deemed to be causally efficacious these collective properties supervene on the properties of individuals. Take the examples above. It is not questioned that fixing the properties of individual members of the mob will also fix the properties of the mob itself, and fixing the properties of individuals constituting the National Bank will fix the properties of the National Bank. Unless there is a difference in individual properties, there will be no difference in collective properties.

So we have a well-developed approach which, by making use of the counterfactual account of causation, promises to make ontological individualism and higher-level property causation compatible. What remains to be seen, finally, is whether responsibility-collectivism fits into this picture. As we have seen, the prospects of responsibility-collectivism depend not just on whether collectives can be causally efficacious, but also on the relationship between the respective causal roles or powers of collective and individual properties. This brings us to the second problem for the responsibility-collectivist. I will argue in the next section that this problem of “too many causes” (Woodward 2008, 250) might well prove intractable for the responsibility-collectivist.

## 6.7 Causal Group-Control and the Problem of Exclusion

Recall that the “programming account” has been proposed not just as a general solution to the exclusion problem, but also to solve an issue specific to the debate about collective responsibility. This specific issue was to find a notion of group-control which does not undermine individual control. Such a notion is thought to be needed because responsibility-allocation was not meant to be a zero-sum game. Robust collective responsibility is supposed to be something more and other than individual responsibility, not a replacement for individual responsibility. As noted, this *desideratum* is only made more pressing by the fact that control exercised by people acting on behalf of collectives seems often to be uncompromised and undiminished.

Accordingly, an additional problem for the responsibility-collectivist is whether a causal understanding of group-control can be made consistent with this approach to robust collective responsibility. This, I will argue in this section, is doubtful. The reason is that none of the available solutions to the problem of “too many causes” is

amicable to responsibility-collectivism, not at least as the position was originally conceived.

There are two ways in which supervenient and subvenient properties could be thought to be related in terms of causal influence.<sup>26</sup> Either one excludes the other. This can happen through downward or upward exclusion. In the case of downward exclusion, the efficacy of the higher-level property entails denying causal efficacy to the lower-level property. In the case of upward exclusion, the opposite process takes place. Alternatively, it is also possible that there is no competition. Supervenient and subvenient properties may both be deemed to be causes of the same effect.<sup>27</sup> Neither of these options is satisfactory for the responsibility-collectivist. Or so I will argue.

I begin with “exclusion scenarios”. Since we are concerned with the question whether properties of collectives, that is, higher-level properties, can be causally efficacious, we can set aside the issue of upward exclusion here. In the case of downward exclusion, the causal efficacy of higher-level properties excludes the causal efficacy of lower-level properties despite the supervenience of the former on the latter. In Lawrence Shapiro’s graphic formulation (2012, 53): “With downwards exclusion, the second-order [i.e., higher-order] property reaches down to snatch all the causal efficacy from its realizer.”

The reason is the multiple realizability of higher-order properties. Given multiple realizability, counterfactuals linking the higher-level supervenient property ( $P_H$ ) to the relevant effect ( $E_H$ ) are both true ( $P_H \square \rightarrow E_H$ ,  $\neg P_H \square \rightarrow \neg E_H$ ),<sup>28</sup> while the negative counterfactual ( $\neg P_L \square \rightarrow \neg E_L$ ) linking the lower-level subvenient property ( $\neg P_L$ ) to the same effect ( $E_L$ ) will be false (List and Menzies 2009).<sup>29</sup> This is because, given multiple realizability, it will not be true that the effect would not have happened without the lower-level property since multiple realizability entails that *other* lower-level properties could have realized the higher-level property. “We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it.” – says Lewis (1973, 557). The point is that due to multiple realizability causal claims about the lower-level property do not satisfy the second part of the conjunct put forward by Lewis. This is because the same effect would have happened without the lower-

---

<sup>26</sup> Many have argued that the exclusion principle should not be treated as an *a priori* claim (see esp. List and Menzies 2009), and that whether exclusion holds or not is to be determined by empirical characteristics of the relevant systems.

<sup>27</sup> But not because they overdetermine the effect. Overdetermination presupposes fully independent property instantiations or events as causes. The underlying rationale of the “non-competition scenario”, as we will see shortly, is that supervenient properties are not wholly distinct from their realizers.

<sup>28</sup> Whereby  $P_H$  and  $E_H$  are propositions referring to the instantiations of higher-level properties, while  $P_L$  and  $E_L$  are propositions referring to the instantiations of lower-level properties.

<sup>29</sup> Diverging from this approach, Woodward (2008) and Shapiro and Sober (2012) argue that even in cases of multiple realizability causal claims involving subvenient lower-level properties need not be false or even less informative.

level property. The absence of the lower-level property does not make a difference when the higher-level property is multiply realizable. Therefore, the higher-level property but not its lower-level realizing property will be causally efficacious.

For example, it is true (other things being equal) that the glass would not have cracked if the water had not boiled, and also that it did crack when the water did boil. But there is a huge number of possible micro states which can realize the relevant value for the temperature of the liquid. So it is not true that had a certain molecular arrangement not broken the glass, the glass would not have cracked. The counterfactual approach entails that the macroscopic, higher-level property is the cause of the cracking, not the microscopic property of a given micro-state of the molecules.

The same reasoning can be applied to collectives. There are many different ways to put together a mob capable and willing to destroy the Bastille. It is very unlikely that the relevant outcome, i.e., the destruction of the Bastille, would be dependent on the exact constitution of the mob. Again, the counterfactual approach implies that the mob *qua* mob was the cause of the Bastille's downfall, and not the properties of individuals which happened to constitute the mob on the given occasion. By the same token, it is highly probable that we will find a robust counterfactual correlation between higher-level property such as the interest rate set by the National Bank's and the relevant outcome such as the level of inflation. That is, "the effect would continue to occur across changes in the lower-level realization of its putative higher-level cause" (List et al. 2012, 17). If so, then the counterfactual account suggests that the setting of the interest rate by the National Bank is the cause of changes in inflation, not the properties of individuals which realize the setting of the interest rate on the given occasion.

I will not question here the argument from multiple realizability to downward exclusion. Rather, the point I want to make here is that if higher-level and lower-level properties are related to one another as in the above scenario of downward exclusion, this will not help the responsibility-collectivist. If the causal efficacy of supervenient properties presupposes the exclusion of subvenient properties, then it would follow that attributing causal efficacy to the collective comes at the expense of denying causal efficacy to individual members of the collective. In short, whenever the collective is a causally efficacious factor, the individual members of the collective cannot be. And since the absence of causal efficacy entails the absence of causal control, it would also follow that causal group-control excludes individual control. Whenever the group is in control in the sense of being causally efficacious, the individual controlled by the collective does not exercise causal control.

The crucial issue is that this understanding is at odds with the responsibility-collectivist's original *desideratum* of finding a notion of group-control which preserves individual control. If causal control is indeed necessary for moral responsibility, this understanding also entails that whenever the group is collectively responsible for some outcome, individual members are not individually responsible for the same outcome. This is because in such cases the causal efficacy attributed to the group regarding that outcome excludes the causal efficacy of individual

members regarding the same outcome.<sup>30</sup> Therefore, contrary to the original intentions of the responsibility-collectivists, what happens is not that collective responsibility will be something over and above the individual responsibility of group-members. Rather, responsibility will shift upwards from the individual level to that of the collective.

This implication, I believe, should seriously worry the responsibility-collectivist. All the more so, if it is true, as is frequently claimed (see, for example, Kincaid 1986; Zahle 2003; List and Menzies 2010), that most collective (and social) properties are indeed multiply realizable. The final point I want to make, however, is that the situation is even less encouraging for the responsibility-collectivist when higher-order and lower-order properties are not in competition for causal efficacy.

Higher-level and lower-level properties need not compete with one another. I now turn to this “non-competition scenario”. The counterfactual account allows for the possibility that both a higher-level property and a lower-level property are causes of the same effect. This is the case when *both* the negative and positive counterfactuals are true for the higher-level property *as well as* for the lower-property realizing it ( $P_H \square \rightarrow E_H$ ,  $\neg P_H \square \rightarrow \neg E_H$ ,  $P_L \square \rightarrow E_L$ ,  $\neg P_L \square \rightarrow \neg E_L$ ). This will clearly be the case when the relevant properties are identical.

Perhaps the higher-level property and the lower-level need not be identical for the same counterfactuals to apply to both of them with regard to some effect. However, and this is the crucial point here, if the same counterfactuals do apply to them with regard to some effect, then it is entailed by the counterfactual account that the higher-level property will have no causal powers or causal role over and above the causal powers or role of the lower-level property. This need not mean that the higher-level property is epiphenomenal, but it does mean that the causal powers or role of the higher-level property is reducible to the causal powers or role of the lower-level property.<sup>31</sup>

It should now be clear why the “non-competition scenario” is also unhelpful for the responsibility-collectivist. If the causal powers or causal role of higher-level properties are nothing over and above the causal powers or causal role of the lower-level properties realizing them, then the causal significance of collective properties too will reduce to the causal significance of the properties of individuals constituting the collective. If that is the case, however, then the collective will have no causal powers and will not play a causal role irreducible to the causal powers or causal role of its individual members. But that entails in turn that collectives cannot be said to exercise causal control. In fact, they cannot even be said to make a causal contribution over and above the causal contributions of their members. And so, in

<sup>30</sup> In addition, Pettit (1996) and List and Pettit (2008) also worry that without individual control it is unlikely that the collective can display rational patterns of behaviour.

<sup>31</sup> This is also admitted by those who think that multiple realizability entails downward exclusion. See, for example, List and Spiekermann (2012, 17): “[...] realization-sensitive causal relations are fully reducible to a lower level of description”.

the “non-competition scenario” there will be no basis for attributing moral responsibility to collectives over and above the moral responsibility of their individual members.

## 6.8 “In Conclusion, Some Ontology”<sup>32</sup>

I summarize the above argument as the responsibility-collectivist’s dilemma. Non-causal group-control does not undermine individual control, but it does not suffice for robust collective responsibility. Causal group-control can suffice for robust collective responsibility (provided other necessary conditions of “responsibility-fitness” have been met), but only in the “exclusion scenario” in which case causal group-control undermines individual control and hence individual responsibility. I have argued that this is bad news for the responsibility-collectivist.

Clearly, some important questions had to be postponed. I have not argued directly that causal control is necessary for moral responsibility. The failure of the “programming account” does provide support for that general claim. As noted, it is also a claim that is intuitively plausible and accepted by most authors on the subject. But not by everyone (see Kutz 2000; Sartorio 2004; Lawson 2012). It remains to be investigated how feasible these challenges to the requirement of causal control really are (for some doubts, see Petersson 2013). This is work for another day.

Instead, in closing I would like to ask about the implications of the conclusions drawn here for the individualism/holism debate. Any lessons learnt with regard to the problem of whether social wholes are something over and above their parts?

We have already noted that ontological individualism, i.e., the property supervenience thesis, is a widely accepted position. It is also assumed by responsibility-collectivists discussed in this paper. However, there are further dimensions of the individualism/holism debate. Pettit and Schweikard highlight three especially significant aspects (using a slightly idiosyncratic terminology, see esp. Pettit and Schweikard 2006). The *individualism* issue, on Pettit’s and Schweikard’s understanding, concerns “the question of whether our individual intentional psychologies are compromised in any way by social regularities” (Pettit and Schweikard 2006, 35; see also Pettit 1996). If individualism in this sense was false, then we would be “predetermined or predestined, notwithstanding our apparent intentional powers, to behave so that the regularities are sustained” (35). The *atomism* issue concerns “the question of whether there are any aspects of our individual intentional psychology such that we depend noncausally on having certain relations with one another for instantiating those features” (35–6). Finally, the *singularism* issue concerns the question of whether there can be group agents in

---

<sup>32</sup> I am importing here a section title from Pettit and Schweikard (2006).

the sense of being “centers of intentional attitude and action over and beyond singular agents” (36).

Clearly, it is the first issue of individualism (in terms of the definition in the previous paragraph<sup>33</sup>) and the third issue of singularism which the conclusions of this paper bear upon. The connection to the third issue is obvious and undisputed by parties to the debate. Responsibility-collectivism presupposes anti-singularism. However, my final point to be made here is that responsibility-collectivism appears to be in tension with individualism. And since responsibility-collectivism presupposes anti-singularism, individualism and anti-singularism also conflict (*pace* Pettit and Schweikard 2006).

Again, the conclusions of this paper do not necessarily clash with the claim that it is possible to ascribe autonomous mental states or attitudes to collectives, i.e., that collectives may have “minds of their own” (Pettit 2003). However, we find that the conditions for responsibility-collectivism are more demanding than its advocates may have previously thought. Specifically, we have seen that collectives can be causally efficacious *qua* collectives only by excluding the causal efficacy of individual contributions. Now, if we combine this conclusion with the “collective mind” hypothesis, we obtain the following. Whenever the collective is a causally efficacious “center of intentional attitude and action” with regard to some outcome, individuals constituting the collective will not be able to control that outcome *via* their intentional attitudes.

And thus, finally, we are in a position to recognize a more general implication, namely that the price of anti-singularism too may be higher than previously thought. This is because it begins to appear questionable whether anti-singularism really is consistent with individualism in Pettit’s sense. If causally efficacious collective agency necessarily compromises individual control in the way described above, then it is hard to see how there can be group agents “distinct in an absolutely clear sense from their own members” (Pettit and Schweikard 2006, 34–5), *while at the same time* the individuals constituting those group agents retain full control over their actions on behalf of the collective.

One way to understand the argument of this paper then is that it diagnoses a tension between individualism and anti-singularism – at least in cases in which the relevant “social regularities” depend on the reality of collective agents. If group agency undercuts the intentional powers of individuals as I have argued, then the price of anti-singularism may well be the denial of individualism. Since the arguments in favour of individualism (persuasively set out in Pettit 1996, among others) appear to be strong, this speaks against anti-singularism in my view. But that polemic is work for another day.

---

<sup>33</sup> The position Pettit and Schweikard call individualism is to be distinguished from responsibility-individualism. The latter view says that only individuals can be the addressees of ascription of responsibility. It is thus opposed to responsibility-collectivism. This section explores the relationship between individualism in Pettit’s and Schweikard’s sense and responsibility-individualism.

## References

- Braham, M., & van Hees, M. (2011). Responsibility voids. *Philosophical Quarterly*, 61, 6–15.
- Copp, D. (2006). On the agency of certain collective entities: An argument from normative autonomy. *Midwest Studies in Philosophy*, XXX, 194–221.
- Crane, T. (2008). Causation, and determinable properties: On the efficacy of colour, shape and size. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced. New essays on reduction, explanation, and causation* (pp. 176–195). Oxford: Oxford University Press.
- Epstein, B. (2009). Ontological individualism reconsidered. *Synthese*, 166, 187–213.
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control. A theory of moral responsibility*. Cambridge: Cambridge University Press.
- French, P. (1979). The corporation as a moral person. *American Philosophical Quarterly*, 16, 207–215.
- French, P. (1984). *Collective and corporate responsibility*. New York: Columbia University Press.
- Gilbert, M. (2002). Collective wrongdoing: Moral and legal responses. *Social Theory and Practice*, 28, 167–187.
- Goetz, S. C. (1988). A noncausal theory of agency. *Philosophy and Phenomenological Research*, 49, 303–316.
- Haji, I. (2006). On the ultimate responsibility of collectives. *Midwest Studies in Philosophy*, XXX, 292–308.
- Held, V. (1970). Can a random collection of individuals be morally responsible? In L. May & S. Hoffman (Eds.), *Collective responsibility* (pp. 89–100). Lanham: Rowman and Littlefield.
- Jackson, F., & Pettit, P. (1990). Program explanation: A general perspective. *Analysis*, 50, 107–117.
- Jackson, F., & Pettit, P. (1992a). In defense of explanatory ecumenism. *Economics and Philosophy*, 8, 1–21.
- Jackson, F., & Pettit, P. (1992b). Structural explanation in social theory. In D. Charles & K. Lennon (Eds.), *Reduction, explanation, and realism* (pp. 97–132). Oxford: Clarendon.
- Kincaid, H. (1986). Reduction, explanation, and individualism. *Philosophy of Science*, 53, 492–513.
- Kutz, C. (2000). *Complicity*. Cambridge: Cambridge University Press.
- Lawson, B. (2012). Individual complicity in collective wrongdoing. *Ethical Theory and Moral Practice*, 16(2), 227–243. Online first.
- Lewis, D. (1973). *Causation* (Philosophical papers, Vol. II, pp. 159–171). New York: Oxford University Press.
- List, C. (2006). The discursive dilemma and public reason. *Ethics*, 116, 362–402.
- List, C., & Menzies, P. (2009). Non-reductive physicalism and the limits of the exclusion principle. *Journal of Philosophy*, 106, 475–502.
- List, C., & Menzies, P. (2010). The causal autonomy of the special sciences. In C. Macdonald & G. Macdonald (Eds.), *Emergence in mind* (pp. 108–128). Oxford: Oxford University Press.
- List, C., & Pettit, P. (2002). The aggregation of sets of judgments: An impossibility result. *Economics and Philosophy*, 18, 89–110.
- List, C., & Pettit, P. (2008). Group agency and supervenience. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced. New essays on reduction, explanation, and causation* (pp. 75–92). Oxford: Oxford University Press.
- List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford: Oxford University Press.
- List, C., & Spiekermann, K. (2012). *Methodological individualism and holism in political science: A reconciliation* (Unpublished Working Paper). <http://personal.lse.ac.uk/list/>
- May, L. (1991). Metaphysical guilt and moral taint. In M. Larry & H. Stacey (Eds.), *Collective responsibility* (pp. 239–254). Lanham: Rowman and Littlefield.
- May, L., & Hoffman, S. (Eds.). (1991). *Collective responsibility*. Lanham: Rowman and Littlefield.

- McKenna, M. (2006). Collective responsibility and an agent meaning theory. *Midwest Studies in Philosophy*, *XXX*, 16–34.
- Menzies, P. (2008). The exclusion problem, the determination relation, and contrastive causation. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced. New essays on reduction, explanation, and causation* (pp. 196–218). Oxford: Oxford University Press.
- Miller, S., & Mäkelä, P. (2005). The collectivist approach to collective moral responsibility. *Metaphilosophy*, *36*, 634–651.
- Moore, S. M. (2009). *Causation and responsibility: An essay in law, morals, and metaphysics*. Oxford: Oxford University Press.
- O'Connor, T. (1994). Emergent properties. *American Philosophical Quarterly*, *31*, 91–104.
- O'Connor, T. (1995). Agent causation. In T. O'Connor (Ed.), *Agents, causes, and events: Essays on indeterminism and free will* (pp. 173–200). Oxford: Oxford University Press.
- Peterson, B. (2013). Co-responsibility and causal involvement. *Philosophia*, *41*, 847–866.
- Pettit, P. (1996 [1993]). *The collective mind*. New York: Oxford University Press.
- Pettit, P. (2003). Groups with minds of their own. In F. Schmitt (Ed.), *Socializing metaphysics* (pp. 167–193). New York: Rowman and Littlefield.
- Pettit, P. (2007a). Responsibility incorporated. *Ethics*, *117*, 171–201.
- Pettit, P. (2007b). Rationality, reasoning and group agency. *Dialectica*, *61*, 495–519.
- Pettit, P., & Schweikard, D. (2006). Joint actions and group agents. *Philosophy of the Social Sciences*, *36*, 18–39.
- Sartorio, C. (2004). How to be responsible for something without causing it. *Philosophical Perspectives*, *18*, 315–336.
- Shapiro, L. A. (2012). Mental manipulations and the problem of causal exclusion. *Australasian Journal of Philosophy*, *90*, 507–524.
- Shapiro, L. A., & Sober, E. (2012). Against proportionality. *Analysis*, *72*, 89–93.
- Shockley, K. (2007). Programming collective control. *Journal of Social Philosophy*, *38*, 442–455.
- Szigeti, A. (2014). Are individualist accounts of collective responsibility morally deficient? In A. Konzelmann-Ziv & H. B. Schmid (Eds.), *Institutions, emotions, and group agents. Contributions to social ontology* (pp. 329–342). Dordrecht: Springer.
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.
- Woodward, J. (2007). Causation with a human face. In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality* (pp. 66–105). Oxford: Oxford University Press.
- Woodward, J. (2008). Mental causation and neural mechanisms. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced. New essays on reduction, explanation, and causation* (pp. 218–262). Oxford: Oxford University Press.
- Yablo, S. (1992). Mental causation. *Philosophical Review*, *101*, 245–280.
- Ylikoski, P. (2012). Micro, macro, and mechanisms. In H. Kincaid (Ed.), *The Oxford handbook of philosophy of social science* (pp. 21–45). Oxford: Oxford University Press.
- Zahle, J. (2003). The individualism-holism debate on intertheoretic reduction and the argument from multiple realization. *Philosophy of the Social Sciences*, *33*, 77–99.
- Zahle, J. (2007). Holism and supervenience. In S. P. Turner & M. W. Risjord (Eds.), *Handbook of the philosophy of science* (pp. 311–341). New York: Elsevier.



# Chapter 7

## Rethinking Micro-Macro Relations

Petri Ylikoski

**Abstract** This paper proposes a new approach to the micro-macro problem in the social sciences. It argues that the common strategy of borrowing arguments from the philosophy of mind debates is not fruitful and the micro-macro relations should not be conceptualized in terms of ‘levels’. This way of thinking is systematically misleading and fails to provide methodologically useful guidance. As a replacement the paper suggests an approach that consider micro-macro relations in terms of scale. In this view there is no unique micro level in the social sciences, and the micro-macro contrast is always context-relative. When combined with the idea of mechanism-based explanation this idea provides an effective tool for thinking about explanation-related controversies in the philosophy of social sciences. For example, by clearly distinguishing causal and constitutive explanations at different scales, it is possible to resolve many conceptual puzzles related to macro causation. The scale-based approach also makes it possible to explore the diversity of macro social properties. To emphasize the importance of this diversity, the paper concludes by presenting a fourfold classification of these properties.

### 7.1 Introduction

The purpose of this paper is to suggest a new way to think about the micro-macro problem (Coleman 1986; Barnes 2001; Bouvier 2011; Molloy et al. 2011) in the social sciences. I argue that the popular approach to this problem, which conceptualizes it in terms of ‘levels’ and borrows concepts and arguments from the philosophy of mind debates, is misguided. It creates pseudo-problems and fails to

---

P. Ylikoski (✉)

Science and Technology Studies/Sociology, Department of Social Research, University of Helsinki, Helsinki, Finland

e-mail: [petri.ylikoski@helsinki.fi](mailto:petri.ylikoski@helsinki.fi)

provide methodologically fruitful guidance. As a replacement for this conceptualization, I suggest an approach that starts from the observation that micro and macro are characterized by different scales. When this idea is combined with the idea of mechanism-based explanation, we have effective conceptual tools for clarifying micro-macro issues facing social scientists.

Philosophers have traditionally assimilated micro-macro issues into the problem of methodological individualism. While the debates about methodological individualism touch real issues, they also carry much unnecessary baggage. An example is the definition of individualism: there are many incompatible definitions of it (and its opposite). Various authors writing about methodological individualism are not actually discussing the same issues, and over the years the term has come to mean many different things (Udéhn 2001; Zahle 2006; Hodgson 2007). Over time, the positions of individualists and their opponents have gotten closer to each other, so that today it is very difficult to distinguish between many current 'individualists' and 'holists' in terms of substantial social scientific claims (Udéhn 2001). As real social scientific examples are rarely used in the debate, it is difficult to see what, if any, practical differences there are between the positions.

Despite this, both the fears and the demands of methodological individualism still haunt debates in the social sciences. This is due to the strong emotions associated with the term 'methodological individualism', which is related to other highly charged but obscure notions like reductionism and political individualism. Furthermore, 'methodological individualism' (without any precise definition) has become a code term for disciplinary identity. In my experience this is particularly the case with economics and sociology. For many economists, a denial of methodological individualism simply signals that a person is confused and does not truly understand economics. In contrast, for many sociologists, supporting methodological individualism signals an anti-sociological reductionist attitude that is bound to overlook many crucial aspects of social reality. Thus the time is ripe for a fresh and less loaded approach to micro-macro relations that leaves behind the old baggage from the methodological individualism debates. The purpose of this paper is to clear the ground for such an approach.

The paper has the following structure. Next in Sect. 7.2 I will argue that the levels-based conception of the micro-macro problem which is borrowed from the philosophy of mind does not really work in the case of the social sciences. Section 7.3 will sketch out an alternative based on the simple idea that macro and micro properties differ from each other by the scale of the entities they are attributed to. Section 7.4 will introduce the idea of mechanism-based explanation and show how it complements the scale approach. Section 7.5 will look at the diversity of social macro properties and will argue that their characteristics are better grasped with scale-based rather than levels-based thinking.

## 7.2 Why Philosophy of Mind Is Not a Good Model

In philosophy of social science, a popular argumentative strategy has been to view the micro-macro problem as analogical to the mainstream way of conceiving the relation between the physical and the mental (Currie 1984; Pettit 1993; Kincaid 1996, 1997; Zahle 2006; Sawyer 2005). Anti-individualists in particular have been excited about the possibility of presenting arguments that are similar to the arguments for non-reductive materialism. The transformation of the philosophy of mind arguments to philosophy of social science arguments is not difficult – quite often philosophers of mind simply talk about M and P predicates. Changing mental to social, and physical to individual (or psychological), is easy, as there is nothing mind-specific about M predicates. The appeal of this strategy is based on the belief that the ideas of supervenience and multiple realization provide a neat way to argue against reductionism. In what follows, I argue that these ideas are less fruitful than commonly assumed and that the philosophy of mind analogy leads us to misconceive the nature of the micro-macro relation.

Let us begin with the notion of supervenience. The basic idea of supervenience is that higher-level properties can be multiply realized by lower-level properties, but once the lower-level properties are fixed, the higher-level properties are fixed as well. As the slogan states, there can be no difference in supervenient properties without a difference in subvenient properties. Over the last 30 years a huge philosophical industry has emerged in connection with this basic idea, and it has been applied to a wide variety of topics including esthetics, metaethics, philosophy of mind, and the philosophies of biology and social science (see Kim 1993; Horgan 1993; McLaughlin 1995; Zahle 2006). The resulting debates point to a whole family of related notions, and philosophers have distinguished literally dozens of different formulations of supervenience. Reviewing these is beyond the scope of this paper, so I will only make some general observations about the applicability of the notion of supervenience to philosophy of social science.

First, the wide applicability of the notion of supervenience suggests that it is a less informative notion than its early proponents believed. If the same notion can be used to characterize the relation between natural and normative facts, esthetic and physical properties, mental and neural states, and biological fitness and its physiological and behavioral bases, it is not metaphysically very informative. What is common among these cases is a kind of property covariance, but this relation is compatible with wide variety of ontological relations (Kim 1993; Horgan 1993). Thus it is no surprise that philosophers of mind have argued that supervenience is compatible with reductive materialism, epiphenomenalism, emergentism, and even straightforward dualism. The lesson for philosophy of social science here is that the notion of supervenience needs to be bundled with substantial ontological ideas in order to be philosophical useful. Thus supervenience alone does not solve major metaphysical problems; on the contrary, its successful application presupposes that one already has a clear ideas of how to face these metaphysical challenges.

Second, while the early enthusiasm over supervenience was based on the vision of non-reductive materialism, the relation between supervenience and reduction has become quite complex (Kim 1993). While the original idea was that supervenience blocks type-identity between higher- and lower-level properties, the situation has turned out to be much more complicated. Whether or not supervenience blocks reduction depends on both how the notion of supervenience is defined and how one understands the notion of reduction. It is therefore no longer viable to assume that supervenience is safe for anti-reductionist arguments, and this means that is much less appealing for philosophers of social science. Further, it is not the case that the concept of supervenience makes it possible to avoid thorny issues related to explanation and causation. Rather, its use in the reductionism debate already presupposes a substantial account of scientific explanation and interfield relations.

The third problem with supervenience in philosophy of social science is that it is methodologically ill-suited. The abstract discussions that relate imagined complete accounts of individual and social levels to each other are far removed from everyday explanatory practice in the social sciences. If the real micro-macro problem is about the relation between small scale and large scale – as I argue in this paper – the use of the notion of supervenience becomes quite difficult. This is because supervenience requires that both the subvenient and the supervenient properties be properties of the same object. In philosophy of mind both mental and neurological properties (or states) are assumed to be attributes of the same object. If one chooses a global or regional form of supervenience one still assumes a parity in size: the subvenient and supervenient attributes occupy the same spatio-temporal area. It is not completely impossible to tweak the notion of supervenience to accommodate the disparity of size between micro and macro – one can assume that the micro basis is whatever underlies the macro property, thus guaranteeing the parity in size. The point is that this artificial stipulation detaches the philosophical discussion from the original methodological challenge the social scientists were facing. Thus it is not an accident that debates about supervenience in philosophy of social science never take real examples from the social sciences.

The problem is not only with the notion of supervenience. My argument is that the underlying mind-brain analogy is an inappropriate model to use when considering the social scientific micro-macro problem. The central problem in philosophy of mind is to understand how the explanations provided by psychological theories that employ mental concepts are related to those provided by the neurosciences. While the vocabularies of these accounts are conceptually discontinuous, they are ultimately talking about the same things. Thus the anti-reductionists do not typically challenge the causal sufficiency of the neural-level facts. For this reason we can say that the philosophy of mind debate presupposes the idea of a comprehensive and exhaustive neural-level understanding of psychological processes.

This is quite different from the social scientific debates about methodological individualism. The problem here is not one of bridging the gap between a comprehensive and exhaustive individual-level understanding of social processes (analogous to the idealized knowledge of the brain) and a more social or holistic description (analogous to the idealized psychological theories employing a mental

vocabulary). Typically the anti-individualists challenge the causal sufficiency of individual facts. They claim that the facts about individuals allowed by the individualist are not sufficient to account for all social facts. Or, they argue that the individualists are cheating by accepting facts that are not properly individualistic. This is because the issue is not actually about the relations between two comprehensive (and potentially competing) levels of description. The challenge that social scientists are facing is one of seeing how local facts about individuals and their social interactions are related to large-scale facts about groups, organizations, and societies.

These observations lead to the conclusion that the analogy between brain-mind (physical-mental) and individual-social does not really work. First, as I already pointed out, these debates have different presuppositions. Second, the ‘individual level’ is not comparable to the ‘physical level’ (Epstein 2009). For example, the former is hardly comprehensive in the way the latter is assumed to be. It is not normally taken to contain things such as artifacts, buildings, microbiological organisms, or household animals, as their inclusion would stretch the idea of individual beyond comprehension. And still these things are often crucial elements of the supervenience basis of macro social properties. For this same reason, the idea of the causal closure of the ‘individual level’ does not make much sense. The third problem is that the ‘social’ is not comparable to the ‘mental’. The social is not really constituted by a vocabulary that is discontinuous with the ‘individual’ level discourse, as in the case of ‘mental’ and ‘physical’. Many macro social properties are aggregates of individual properties, some macro properties are similar to micro properties, and some are anthropomorphic projections of personal properties (for example, the attribution of beliefs and desires to groups and organizations). Clearly, there is room for a fresh start.

### 7.3 The Scale-Based Alternative

If the notion of supervenience is of doubtful value and the analogy with the philosophy of mind is misleading, it makes sense to rethink the issue. This is what I will do next. I will begin with some general observations about macro social facts and then suggest an alternative to the philosophy of mind-inspired account. This new account leaves out the metaphor of levels of reality (see Oppenheim and Putnam 1958; Kim 2002) and considers micro-macro relations as issues of scale (Ylikoski 2012). I will argue that this new approach is more sensitive to the real micro-macro problems social scientists are facing and that it avoids many philosophical problems generated by the imagery of levels. Let us start with the general observations.

First is the observation that macro social facts are typically supra-individual. Social macro properties are attributed to groups, communities, populations, and organizations. Some attributes apply to both individuals and collectives, but typically macro social properties, relations, and events are not about individuals.

A second salient feature is that micro and macro have a part-whole relationship: the macro social entities are made up of their constituting parts. This relation is not a mere mereological aggregation or simple material constitution, because the relations between the components play an important role. Furthermore, the relations between social wholes, and more generally between the social whole and its environment, are often of crucial importance. In addition, most social wholes are composed of a heterogeneous set of entities, usually intentional agents, their ideas, and material artifacts.

These points make it possible to see the micro-macro relation as a question of scale: the difference between micro and macro is the difference between small- and large-scale social phenomena. The point here is not to simply define the micro-macro contrast as an issue of scale. After all, most differences in scale do not constitute a meaningful micro-macro relation, and the heterogeneous nature of macro social facts makes it difficult to characterize the additional requirements for their defining features. What I want to suggest is that the scale perspective provides a fruitful heuristic way of thinking about micro-macro relations that avoids many problems associated with the levels view.

A consequence of this approach is that it permits the idea that there is *no unique micro level in the social sciences*. The traditional contrast between ‘individual’ and ‘social’ levels is categorical, but the contrast between small and large scale is relative. The micro and macro can be of different sizes: depending on the application, the micro entities could be individuals, families, firms, or groups. This is precisely the way social scientists talk about micro-macro relations. They are quite flexible about the size of their micro and macro properties, and they do not assume that micro is always about one specific set of entities. Thus there is no need to postulate a specific ‘individual level’ to serve as a general micro level for all purposes. When the idea of ‘level’ is given up, we can also give up the idea of a comprehensive and privileged level of explanation (Ylikoski 2012).

Social scientists also regard the micro-macro contrast as context-relative: whether an attribute is a macro or micro property depends on what it is contrasted with. A friendship relationship is a macro property from a psychological point of view, but a micro property when considered from the point of view of the social networks within a community. According to the scale view, the contrast between micro and macro depends on one’s explanatory interests, not on a priori considerations. For example, international politics and organizational sociology construct the micro-macro contrast quite differently. In the former, states and other organizations are often treated as individuals, whereas in the latter, the organizations and their characteristics are the macro reality to be explained. Similarly, an economist studying market processes can treat firms and households as the micro elements, while in industrial organization and family sociology, these are often the macro items to be explained.

The scale-based view can also accommodate the idea of meso level that sometimes emerges in social scientific discussions (Jepperson and Meyer 2011; Little 2012). Often it makes sense to analyze social processes at some intermediate scale. In fact, in the mechanism-based account this is what one should do: track how

causal influences are transmitted between processes at different scales. However, in contrast to the levels view, one does not have to assume that there is somehow a fixed meso level, nor is there any need to constantly come up with new names for new levels (for example, for the level between micro and meso, or between macro and meso). There is a continuum of scale between any specified micro and macro scale, and it is an empirical issue which intermediate scale is most crucial for understanding the micro-macro relation, not something that is set by a priori ontological considerations.

## 7.4 Explanatory Confusion with Levels

The change in guiding metaphors can only take us this far, so we also need an account of explanation. In recent years I have been advocating a mechanism-based approach to social scientific theorizing (Hedström and Ylikoski 2010; Ylikoski 2011, 2012, see also Hedström 2005) that is based on the contrastive-counterfactual account of explanation (Ylikoski 2001, 2007, 2011; Ylikoski and Kuorikoski 2010, see also Woodward 2003). I will not repeat here what I have argued elsewhere, but will point out how this approach can be combined with scale-based thinking to clarify issues related to the explanatory relevance of macro properties.

What is crucial from the point of view of my argument is the distinction between causal and constitutive explanation that has mainly been discussed in philosophy of biology (Craver 2007). This distinction is very important for figuring out what goes wrong with typical micro-reductionist arguments. I have discussed the distinction more systematically elsewhere (Ylikoski 2012, 2013), so here I will be brief.

In the contrastive counterfactual account of explanation the purpose of explanation is to track relations of dependence. These dependencies can be causal, constitutive, or formal. Let us forget the last, and focus on causal and constitutive relations. Causation is usually regarded as a relation between events; it is about changes in the properties of things. A temporal difference between causes and effects is also typically assumed. In other words, causation takes time and thus it is natural to talk about causal processes. These characteristics can be contrasted with the characteristic features of constitutive relations. First, constitution relates properties, not events. It is a relation between a system's properties and the properties of its parts and their organization. We could say that the whole (the system) is comprised of its parts and their relations. In contrast to causation, it makes no sense to think about this relation as a process. A glass is fragile because it has a particular molecular structure: the molecular structure of glass constitutes its fragility. This takes no time, and the fragility is not a product of a process that starts with the molecular structure. Thus having the specific molecular structure and being fragile cannot truly be regarded as 'independent existences'. For this reason the asymmetry of manipulation (Woodward 2003) that characterizes causation cannot be used to characterize constitution. (For a more extensive discussion, see Ylikoski 2013.)

As the point of explanation is to track the relations of dependence, we have a way to distinguish between causal and constitutive explanations. Causal explanations track causal dependencies (Woodward 2003) and constitutive explanations track constitutive dependencies (Craver 2007; Ylikoski 2013). Causal explanation tells us how the antecedent events and their organization (timing and location) brought about the event to be explained. Constitutive explanation describes how the properties of the components and their organization give rise to the causal capacities of the system. I have argued elsewhere (Ylikoski 2013) that despite their subtle metaphysical differences, the principles that guide these explanations are surprisingly similar. Both explanations attempt to track networks of counterfactual dependence, and their pragmatics are very similar. These similarities probably explain why people so often miss the difference between causation and constitution.

The distinction between causation and constitution also applies in the social sciences (Vromen 2010).<sup>1</sup> Consider two groups of equal size that differ in their problem-solving capacities. What could explain this difference in their causal capacities? We can ask either “what makes this group have these properties?” (the constitutive question) or “how did this group acquire these properties?” (the causal question). An answer to the causal question would have examined the group’s and its members’s past, looking for some crucial difference makers. In contrast, the answer to the constitutive question would have looked at things that the group was made of. The crucial difference might be in the properties of the group’s members, such as their intelligence or social skills. Alternatively, the crucial factors might be the informal social norms that characterize the interactions within the group or its formal organization. The first suggestion would appeal to the properties of the system’s (the group) parts, while the latter would refer to their organization. Both of these explanations, and their combinations, would be constitutive explanations.

The distinction between causal and constitutive questions can be made with respect to all social macro properties. The constitutive questions ask how the macro properties of the whole (a group, a population, an organization, or a society) are constituted by the smaller-scale entities, activities, and relations. The aim is to understand how the details of macro-scale facts depend on the micro-scale facts and their organizations. The issue can be captured by asking how the macro facts would have been different if some of the micro facts had been different in some specific way. In contrast, causal questions about the macro social properties concern their

---

<sup>1</sup> Alexander Wendt (1998, 1999: 77–88) is an early advocate in the social sciences of the idea of constitutive explanation. The problem with his discussion is that, while his notion of constitution contains the idea of constitution of causal capacities, it also contains many other ideas. For example, he fails to see that many how-possibly explanations are causal: they are about necessary conditions for something to happen. He also confuses the criteria for category membership with constitutive relations: asking what makes an entity a member of a certain kind is quite different from asking what makes it have a certain property. Finally, he associates the contrast between causation and constitution with that explanation and understanding in a manner that is ultimately unhelpful.



origin, persistence, and change. Here the focus is on tracking counterfactual dependencies between events. The issue is what the outcome would have been if some of the things in the causal history had been different in some specific manner.

Having dealt with these preliminaries, we can now get to the main points. The *explanantia* of constitutive explanations are always at the micro scale. As the explanation attempts to capture what the whole is made of, an appeal to the properties of the whole does not make much sense. One could say that in this sense the methodological individualists, and other reductionists, have been on the right track in their claims about the explanation of social macro properties. However, as they have not recognized the differences between causal and constitutive explanations, the micro-reductionists have been led to make some unwise claims about the causal explanation of social macro properties. The notion of constitution implies that every time we have a causal capacity attributed to a macro-scale entity, we also have micro-scale facts that constitute it. However, this does not imply that causal explanations should be micro-based. This would require a separate argument that is not usually provided.

The constitutive explanation of macro properties does not in any way diminish their reality or explanatory relevance in causal contexts: the wholes are as real as their parts. To say that micro explanations somehow eliminate the macro properties is simply metaphysically confused. This confusion is suggested by the metaphor of levels, which makes it possible to think of micro and macro properties as competing causes. But if the causal capacity of the whole and its micro basis are not distinct existences, this situation is not possible. There is only one thing. One could say that all causation is at the same level, but this makes little sense, as the idea of a level requires a contrast. What really matters is the scale: examples from physics, biology, and human sciences all show that it matters how things are organized, and this implies that all explanatory factors are not found at the most local micro scale.

When anti-reductionists talk about emergence (Sawyer 2005; Elder-Vass 2010), they usually want to emphasize the importance of organization in the explanation of complex processes. However, as the idea of emergence is articulated within the context of ‘levels’, it often leads to quite strange results (McLaughlin 1992; Kim 1999). For example, there has been good deal of metaphorical discussion about downward causation.<sup>2</sup> Sometimes this idea sounds like something Baron Münchhausen could have invented, and sometimes it seems to be a complicated way of expressing quite uncontroversial ideas (Kim 1999; Robinson 2005; Craver and Bechtel 2007). However, from the point of view of the theory of explanation,

---

<sup>2</sup> The old philosophy of social sciences debates about the possibility of macro-level laws overriding individual agency (*autarchy*) (see Zahle 2006; Pettit 1993) are basically special cases of the downward causation issue with some added elements about the freedom of human action. The theme of freedom and determinism is also at the core of the agency/structure debate, which can be characterized as a sociological version of the problem of free will (Loyal and Barnes 2001) although some real empirical issues are also related to this theme, e.g. the opportunities open to an agent, and an agent’s beliefs about her opportunities (Hitlin and Long 2009).

the ultimate problem with the concept of emergence is that it merely names a class of phenomena without providing any tools to analyze the role that organization plays in explanation (Kuorikoski and Ylikoski 2013). It is an advantage of scale-based thinking that it allows us to clear away the confusion that the notion of emergence has given rise to, without losing the real issues to which it is attached.

The micro-reductionists are often driven by the intuition that “all real causal work is done at the physical level”, but such intuition is not an argument. It is only a confused idea. This is because people who appeal to this intuition are physicalists, who think that everything is ultimately made of the stuff studied by physics. In other words, everything is constituted by the physical. But if the *relata* in constitutive relations are not distinct existences, it does not make sense to say that one is more real than the other. In fact, nothing in the idea of causation itself would require a regress to the micro level. The same point applies to the social science version of such intuition which says that whatever real happens is at the level of individuals. Even if the idea of ‘individual level’ could be made precise, the intuition is based on the same fallacy about constitution. In the counterfactual account of causal relevance (Ylikoski 2001; Woodward 2003; Steel 2006), the location of explanatory relevance at the micro or macro level is a contingent matter that depends on the *explananda* that one is addressing. There is no a priori reason to assume that the most invariant counterfactual dependences (with respect to the specified *explanandum*) will always be found at the micro level (Woodward 2003, 2008). In this scheme both small- and large-scale things can be crucial difference makers. It is sufficient that there is an appropriate counterfactual dependence between the macro variable and the *explanandum*. Thus the ‘deep’ philosophical problem of macro causation is replaced with the more down-to-earth problem of explanatory selection: under which description can we formulate the most robust claims about counterfactual dependence? The issues of explanatory relevance (how the explanatory factors are selected, at which level of abstraction they are described, etc.) are determined by the facts of the case and the details of the intended *explanandum*, not by generic philosophical arguments.

The contrastive-counterfactual account of explanation also provides a natural antidote for another harmful bias associated with levels-based thinking. This is the anti-reductionist presumption that causes are always at the same level as the effects (‘macro effects have macro causes’). This idea is as wrong and as confused as the micro-reductionist idea that real causes are at the lower level (Woodward 2008). There is no reason to assume that causes are always of the same size as the effects. The causal relevance of things should be decided on an empirical basis, and not be based on philosophical prejudice. Again, this is a biased intuition that is founded on the metaphor of levels. The idea’s appeal disappears once one begins to think of the issue as a matter of scale. I would claim that it is a major advantage of the scale approach that it helps to mitigate the influence of confused intuitions like this on theorizing.

Another advantage of this approach is that allows us to characterize individualism by its real research heuristics rather than hiding them by abstract philosophical talk. What people, rightly or wrongly find disturbing in real-life examples of

methodological individualist theorizing is the dominant focus on individuals and their local interactions. (Compare this to the reductionist research strategies discussed by Wimsatt 2007.) Critics think that these approaches might miss the causal importance of larger-scale structural factors. This point is simply lost if the issues are viewed in terms of comprehensive micro and macro levels. Similarly, in levels-based thinking the problem with holists seems to be their mysterious metaphysical beliefs about ‘downward causation’, while their real problems might be much more mundane. For example, they might be making overly strong assumptions about the stability and relevance of structural facts, or they might be failing to pay enough attention to how their ‘structural’ facts are actually constituted. Again, levels-based thinking hides rather than helps to resolve the real theoretical and empirical issues related to micro-macro relations.

In the scale-based approach, explanatory questions divide into two general classes: causal and constitutive. Causal questions ask: ‘How do the macro-scale changes influence micro-scale processes?’ ‘How do the micro-scale changes influence macro-scale processes?’ In both cases we are interested in the mechanisms by which these influences are transmitted. Constitutive questions, for their part, ask ‘How do the macro-scale things have their causal capacities?’ In other words, they ask how they are constituted by micro-scale entities, processes, and relations. The scale approach to the micro-macro issue suggests a kind of flat view of society in which the difference between micro and macro is one of scale, not of different levels. The large-scale social facts have an irreducible explanatory contribution to make, but there is nothing comparable to the mind-brain relation. In this view, the world is not imaged as a layered cake but as a flat surface. Such a change in metaphors might sound arrogantly reductionist to some, but the above discussion shows that this is a false impression. The scale-based approach provides a fruitful way to formulate micro-macro issues in an empirically tractable manner and to diagnose what is wrong in the traditional macro causation debate.

## 7.5 The Diversity of Social Macro Properties

One of the implicit assumptions of the traditional debate is that social macro properties are homogenous. When the debate is conducted at the abstract level of S and I predicates and without real social scientific examples, this may go easily unnoticed. For the same reason, people have not recognized that social macro facts are actually quite difficult to think of in terms of ‘higher’ and ‘lower’ levels. In this Section I describe how the scale-based approach could naturally replace the metaphorical -levels discussion.

The diversity of social macro properties is a quite unexplored area in philosophy of the social sciences. We do not have a satisfactory classification of different sorts of social macro properties. However, it is useful to have some sort of preliminary taxonomy of macro social facts. Thus I employ the classification proposed in Ylikoski (2012). While it is still an open issue whether this scheme is ultimately

satisfactory, it can be used to demonstrate the heterogeneity of social macro properties and to make the case for the relative fruitfulness of the scale-based approach in analyzing their role in explanations.

Ylikoski (2012) classifies social macro properties into four classes: (1) the statistical properties of a population; (2) the networks of relations within a population; (3) the communal properties; and (4) the properties of organizations. In what follows, I will first briefly describe each case and consider how cogent the competing conceptualizations actually are in capturing their specifics.

When methodological individualists discuss social macro facts, they often have in mind the *statistical properties of a population* (Ylikoski 2012: 28–29). These describe the distributions and frequencies of properties within a given population. The facts about distributions describe how certain attributes are distributed among members of a population or how individuals with certain attributes are distributed among social positions and spatial locations. The facts about frequencies specify the attributes, behaviors, or beliefs that are typical, rare, dominant, or marginal within the population. These properties seem rather unproblematic. They are based on data that may be from pre-existing registers or which are derived from surveys. The data tell about the individual attributes, and the macro properties are inferred (or estimated) from these data points. Thus they are logically continuous with the individual attributes (Pettit 1993: 121). It is no wonder that individualists do not usually find these properties to be threatening. The only problem is with the individual properties themselves. For example, unemployment is an institutional status that cannot be understood as an intrinsic property of an individual. However, we can pass over this problem in the context here.

Is it possible to think about statistical properties in terms higher and lower levels? The first thing to note is that because statistical properties are in principle based only on an aggregation of individual properties, it would be easy to think of them as ‘individual-level’ properties. The problem is that while this stipulation gives an easy (partial) victory to the traditional individualist, it misses the contrast between micro and macro. Thus all micro-macro problems simply become invisible. The alternative is to take them to be macro properties. However, because these properties are logically continuous with the individual attributes, the individualist must invoke something like the scale of the phenomenon. While the individual data points are about individuals, it makes no sense to attribute the statistical properties of populations to individuals. These are facts about the population. For example, the rate of unemployment is an attribute of a specified population, not of any of its individual members. Given that a population consists of its members, the statistical properties of the population are attributed to an entity of a different scale than that of the individual. This clearly demonstrates that the scale-based approach is unavoidable for statistical macro properties. The remaining question is whether the idea of levels adds anything useful to it.

The answer seems to be negative. The first problem is that statistical units do not have to be individuals, they can as well be families or companies, for example. The statistical tools can be flexibly used with units of various scales, and therefore they are not in any way anchored to a specific ‘individual’ level. Similarly, the size of the

population varies, and one population can be a subpopulation of another, overlap, or be completely separate. The size of the entity that has the statistical properties is quite variable. These issues are handled by the scale-based approach quite naturally, but are difficult for thinking based on the metaphor of levels.

The explanatory role of statistical properties provides more challenges to the levels view. While it is easy to think that statistical properties are merely summaries of individual properties and thus statistical causal claims are merely summaries of individual causal claims, this is not the case. Consider the case of frequency-dependent causation. In frequency-dependent causation the causal effect of an individual having a certain property depends on the frequency of that property in the population. While in cases like this there is always something about the individual that makes it sensitive to the frequency of the attribute in the population, frequency still has a crucial and non-reducible explanatory role. It is a crucial difference maker. Within the levels approach this is difficult to capture: In what sense is the frequency of the property in the population a higher-level property that has a downward influence on the individual? It is much more natural to conceive of this as a case where a larger-scale fact has an influence on facts on the smaller scale.

The *network properties of a population* are the next class of social macro properties to consider (Ylikoski 2012: 29–30). These are based on the relations and interactions between individuals. When we have a group of related individuals, we have a network of social relations within that population of individuals. A social network such as this can be regarded as a map of the relevant ties between the group members. The analysis of social networks has become a major interdisciplinary research topic involving sociologists, economists, physicists, and mathematicians. Modern network analysis is based on the observation that networks have many interesting (formal) properties, such as centralization, cohesion, density, and structural cohesion (Newman 2010; Kadushin 2011).

The social network is inferred from knowledge about individual relationships, but the properties of the network are prototypical macro properties. As with statistical properties, it does not make sense to attribute the properties of the whole network to the network's individual nodes. Similarly to statistical properties, the units of network analysis are flexible. There is no requirement that the nodes of the network (the members of the population) be persons. They can be groups, families, organizations, or even states. The vocabulary of levels has difficulty capturing this flexibility, while it is easily captured with scale-based thinking.

As an example, consider the notion of a structural hole (Burt 1992), which can be used to explain the differences in the abilities of agents to access information and their opportunities to influence social processes. In explanations appealing to the existence of structural holes, the structure of the network plays an irreducible role, and it is quite natural to think of the social network as a large-scale social phenomenon influencing local interactions between individuals. In contrast, it is very difficult to see this in terms of social and individual levels, not to mention 'higher' and 'lower' ones. As social networks are attributes of the population, it would be far-fetched to call social networks individual properties, and this stipulation would lead to problems that are similar to those faced with statistical

properties. On the other hand, if we regard network properties as higher-level macro properties, what would be the individual-level properties that could be regarded as their bases? Collections of relevant individual relations, one might answer, but that would be a vague way to talk about networks. Thus it is safe to conclude that the scale-based approach has an advantage over levels-based thinking when it comes to network properties.

The third case to be discussed is *communal properties* (Ylikoski 2012: 30–31). This general term covers a variety of social scientific notions that are applied to specific communities, but not to isolated individuals. Things like culture, customs, and social norms are good examples of communal properties. Social norms and customs are properties of communities – attributing them to solitary individuals does not make sense. Similarly, cultural differences are primarily found between groups, not individuals. Even if communal properties are attributed to groups, they are quite straightforwardly based on facts about individuals. Underlying all these notions is the idea that the members of a group share certain beliefs, expectations, preferences, habits, etc.

What is crucial for communal properties is that the individuals do not share these attributes accidentally: the members have the relevant individual properties because the other members of the group have them. For example, the existence of a social custom presupposes that the novices learn specific expectations and habits when they become members, and that the members of the group maintain these expectations and habits because others do so also. Thus the members share the relevant properties due to a continuing interaction with each other. Similarly, the cohesion of a culture is based on the frequency of interactions with the group and the rarity of interactions with outsiders, not some kind of ‘higher-level’ influence on individuals.

Descriptions of customs, social norms, and cultures are always based on idealization and abstraction. Members of a community never have exactly the same ideas, preferences, or routines. That would be a miracle, given what is known about human learning and communication (Sperber 1996). There is always some variation among the members, no matter how comprehensive the socialization processes are. However, these idealized descriptions are useful. They draw attention to features of the group that are typical and salient when it is contrasted with some other group. Usually the descriptions of larger-scale communal properties are more abstract and less rich in detail, as individual variation takes its toll. In this context it is natural to talk about *levels of abstraction*.

The important point here is that the idea of level of abstraction is quite different from the idea of level of reality, which is the target of my criticism. Consider two descriptions. The first says that certain social norms are in force in a group and the second says the members of the group have certain beliefs, expectations, and dispositions to sanction certain behaviors. These two descriptions do not mention separate facts occupying different layers of reality. Rather, they are two ways of describing the same thing. The norm description is more abstract. It ignores many individual variations between the group members and sacrifices some of the details. The point of using this description is that it captures a salient and relatively stable

feature of the group. Such abstract description is ‘multiply realizable’, i.e. it is compatible with multiple configurations of individual attributes. This kind of abstraction can be useful for some explanatory purposes, but fails for others. For example, when we are explaining the behavior of an individual by appealing to social norms, we are referring to larger-scale facts about the group members that are causally relevant to the micro-level behavior. There is no need to postulate a separate realm of norms to understand what is happening. It is just that the expectations and responses of the other group members influence the individual’s judgments about appropriate behavior. Whether the more abstract description provides a better explanation is an empirical issue about the range invariance of counterfactual dependencies (Woodward 2003; Steel 2006), not an issue that is decided by metaphysical argumentation. Therefore it makes no sense to argue against communal properties like social norms by appealing to the causal impotence of macro-level causes (contra Jones 2010).

Communal properties are tied to a social community defined by frequent interactions, but there is no unique way to identify the relevant community. This makes it possible, for example, to describe culture at various scales, such as a village, a local area, and a nation. The same applies to many other communal properties. As well, favoring the scale-based approach makes it possible to attribute communal properties to non-personal units. For example, it is possible to describe the social norms that govern interactions between organizations. Again, it seems that the metaphor of levels adds nothing useful to the considerations of scale.

The last case to consider is *properties of organizations* (Ylikoski 2012: 31–32). Organizations such as states, firms, parties, churches, and sport clubs are important parts of the modern social reality. In contrast to loosely defined communities that carry communal properties, organizations usually have specified criteria for membership, at least for their operational members. Organizations are based on (written or non-written) rules. They define the rights and duties of members and the roles of various functionaries. Because of the rules, the organizations can remain stable and continuous even if their members change.

Organizations are ontologically interesting because they are entities that can have many properties that are not those of their members. For example, they may have goals that are not the personal goals of their members, and some organizations are even treated as legal persons. However, organizations are ultimately human artifacts made of people and their ideas about the rules, and often material artifacts as well. Whatever causal capacities the organization has, they are constituted by the latter sort of facts. Whatever the organization does, it is done by its members in its name. It is often of crucial social importance whether an action – for example a questionable comment – was made as representing an organization or a private person. However, this is a question about the status attributed to the behavior, not about the causal capacities of two completely separate entities.

While individual organizations are historical entities, their causal capacities are constituted by the causal capacities of their parts and their organization. The organization can be regarded as a system that may have as its parts smaller organizations, individual human beings (having specified beliefs, expectations,

habits, etc.), or technological artifacts (such as texts, computers, means of communication etc.). Finally, we have a case where we could talk about system-level macro properties and their micro-bases. However, even here we can question whether levels-based thinking is really helpful. Consider the causal interaction between an individual and an organization. Within the levels-based approach this raises a spectrum of philosophical problems: Can the higher-level property of an organization be causally relevant if it is fully constituted by micro-level individual properties? More specifically, if the individual is a part of the organization, how can the latter causally influence one of its own parts? From the scale-based point of view these questions are an unnecessary nuisance. When a person causally interacts with an organization, she interacts with other parts of it.

The influence of the organization on its members occurs through other members, no matter how high up some of them are in the organizational hierarchy. There is no downward causal influence from a higher level.

These observations suggest that even in the case of organizations, the levels-based layer-cake model of the social world is not very illuminating. What is interesting about organizations are the habits and mental representations of their members, the resources they control as members of the organization, and their (materially mediated) interactions. While the imagery of levels leads to metaphysical considerations that do not help in testing and formulating substantial social scientific theories, the scale-based approach makes it possible to formulate micro-macro problems in an empirically tractable manner. Instead of metaphysical considerations, one should ask how large-scale collective enterprises – like organizations – manage (or fail) to achieve certain things and how these collective enterprises influence their members. The answers to these questions will often refer to organizations and their properties, but it is not problematic to conceive them as large-scale things influencing smaller-scale things or other large-scale things.

The above discussion suggests that philosophers of social science should turn to scale-based thinking if they want to tackle micro-macro problems in the social sciences. Real examples from the social sciences can be used to replace abstract (and ill-formulated) metaphysical considerations about causal potency and impotence. Furthermore, the diversity of social macro facts suggests that we should not expect a general philosophical solution to micro-macro problems in the social sciences. This makes the case study approach even more appropriate.

## 7.6 Conclusion

In this paper I have advocated the scale-based approach to micro-macro problems in the social sciences. The main claim is that it captures what is puzzling about the micro-macro relations better than the traditional model based on the idea of levels. When combined with the mechanism-based approach to explanation, it provides a fruitful way to address micro-macro issues. The target of my criticism has been the



powerful metaphor of levels that postulates social reality to consist of comprehensive and fixed levels. My suggestion is that once we give up this metaphor, the discussion of micro-macro issues will become much more productive. Of course, many other ways of talking about ‘levels’ are not considered by my critique. Some of these ideas might be legitimate, and so I do not advocate a comprehensive purging of levels-talk from the social sciences. However, I hope I have made the case for discarding one popular way of talking about levels and for a more general requirement that people be clear what they mean when they employ the metaphor of level in the social sciences.

## References

- Barnes, B. (2001). The macro/micro problem and the problem of structure and agency. In G. Ritzer & B. Smart (Eds.), *Handbook of social theory* (pp. 339–352). New York: SAGE.
- Bouvier, A. (2011). Individualism, collective agency and “micro-macro relation”. In I. C. Jarvie & J. P. Zamora-Bonilla (Eds.), *The SAGE handbook of the philosophy of the social sciences* (pp. 199–216). New York: SAGE.
- Burt, R. (1992). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.
- Coleman, J. S. (1986). Social theory, social research, and a theory of action. *American Journal of Sociology*, *91*, 1309–1335.
- Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. New York/Oxford: Clarendon Press.
- Craver, C., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, *22*, 547–563.
- Currie, G. (1984). Individualism and global supervenience. *British Journal for the Philosophy of Science*, *35*, 345–358.
- Elder-Vass, D. (2010). *The causal power of social structures. Emergence, structure and agency*. Cambridge: Cambridge University Press.
- Epstein, B. (2009). Ontological individualism reconsidered. *Synthese*, *166*, 187–213.
- Hedström, P. (2005). *Dissecting the social. On the principles of analytical sociology*. Cambridge: Cambridge University Press.
- Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology*, *36*, 49–67.
- Hitlin, S., & Long, C. (2009). Agency as a sociological variable: A preliminary model of individuals, situations, and the life course. *Sociology Compass*, *3*(1), 137–160.
- Hodgson, G. M. (2007). Meanings of methodological individualism. *Journal of Economic Methodology*, *14*(2), 211–226.
- Horgan, T. (1993). From supervenience to superdupervenience: Meeting the demands of a material world. *Mind*, *102*, 555–586.
- Jepperson, R., & Meyer, J. W. (2011). Multiple levels of analysis and the limitations of methodological individualisms. *Sociological Theory*, *29*(1), 54–73.
- Jones, T. (2010). Norms and customs: Causally important or causally impotent? *Philosophy of the Social Sciences*, *40*(3), 399–432.
- Kadushin, C. (2011). *Understanding social networks: Theories, concepts, and findings*. Oxford: Oxford University Press.
- Kim, J. (1993). *Supervenience and mind*. Cambridge: Cambridge University Press.
- Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, *95*, 3–36.

- Kim, J. (2002). The layered model: Metaphysical considerations. *Philosophical Explorations*, 5, 2–20.
- Kincaid, H. (1996). *Philosophical foundations of the social sciences. Analyzing controversies in social research*. Cambridge: Cambridge University Press.
- Kincaid, H. (1997). *Individualism and the unity of science. Essays on reduction, explanation, and the special sciences*. Lanham: Rowman & Littlefield.
- Kuorikoski, J., & Ylikoski, P. (2013). How organization explains. In D. Dieks & V. Karakostas (Eds.), *EPSA11 perspectives and foundational problems in philosophy of science* (pp. 69–80). Cham: Springer.
- Little, D. (2012). Explanatory autonomy and Coleman’s boat. *Theoria*, 74, 137–151.
- Loyal, S., & Barnes, B. (2001). “Agency” as a red herring in social theory. *Philosophy of the Social Sciences*, 31, 507–524.
- McLaughlin, B. (1992). The rise and fall of British emergentism. In A. Beckermann, H. Flohr, & J. Kim (Eds.), *Emergence or reduction? Essays on the prospects of nonreductive physicalism* (pp. 49–93). Berlin: Walter de Gruyter.
- McLaughlin, B. (1995). Varieties of supervenience. In E. E. Savellos & U. D. Yalçin (Eds.), *Supervenience. New essays* (pp. 16–59). Cambridge: Cambridge University Press.
- Molloy, J. C., Ployhart, R. E., & Wright, P. M. (2011). The myth of “the” micro-macro divide: Bridging system-level and disciplinary divides. *Journal of Management*, 37(2), 581–609.
- Newman, M. E. J. (2010). *Networks. An introduction*. Oxford: Oxford University Press.
- Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Concepts, theories, and the mind–body problem, Minnesota studies in the philosophy of science II* (pp. 3–36). Minneapolis: University of Minnesota Press.
- Pettit, P. (1993). *The common mind. An essay on psychology, society, and politics*. Oxford: Oxford University Press.
- Robinson, W. S. (2005). Zooming in on downward causation. *Biology and Philosophy*, 20, 117–136.
- Sawyer, R. K. (2005). *Social emergence. Societies as complex systems*. Cambridge: Cambridge University Press.
- Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Oxford: Blackwell.
- Steel, D. (2006). Methodological individualism, explanation, and invariance. *Philosophy of the Social Sciences*, 36, 440–463.
- Udehn, L. (2001). *Methodological individualism: Background, history, and meaning*. London: Routledge.
- Vromen, J. (2010). MICRO-foundations in strategic management: Squaring Coleman’s diagram. *Erkenntnis*, 73(3), 365–383.
- Wendt, A. (1998). On constitution and causation in international relations. *Review of International Studies*, 24, 101–117.
- Wendt, A. (1999). *Social theory of international politics*. Cambridge: Cambridge University Press.
- Wimsatt, W. (2007). *Re-engineering philosophy for limited beings. Piecewise approximations to reality*. Cambridge, MA: Harvard University Press.
- Woodward, J. (2003). *Making things happen. A theory of causal explanation*. Oxford: Oxford University Press.
- Woodward, J. (2008). Mental causation and neural mechanisms. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced. Essays on reduction, explanation, and causation* (pp. 218–262). Oxford: Oxford University Press.
- Ylikoski, P. (2001, May). *Understanding interests and causal explanation*. Ph.D. thesis, University of Helsinki. Available at: <http://ethesis.helsinki.fi/julkaisut/val/kayta/vk/ylikoski>
- Ylikoski, P. (2007). The idea of contrastive explanandum. In J. Persson & P. Ylikoski (Eds.), *Rethinking explanation* (pp. 27–42). Dordrecht: Springer.
- Ylikoski, P. (2011). Social mechanisms and explanatory relevance. In P. Demeulenaere (Ed.), *From social mechanisms to analytical sociology* (pp. 154–172). Cambridge: Cambridge University Press.

- Ylikoski, P. (2012). Micro, macro, and mechanisms. In H. Kincaid (Ed.), *The Oxford handbook of philosophy of the social sciences* (pp. 21–45). Oxford: Oxford University Press.
- Ylikoski, P. (2013). Causation and constitution compared. *Erkenntnis*, 78, 277–297.
- Ylikoski, P., & Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical Studies*, 148, 201–219.
- Zahle, J. (2006). Holism and supervenience. In S. P. Turner & M. Risjord (Eds.), *The handbook of philosophy of science. Philosophy of anthropology and sociology* (pp. 311–341). Amsterdam: Elsevier.

**Part II**  
**Methodological Individualism-Holism**

# Chapter 8

## Dead Ends and Live Issues in the Individualism-Holism Debate

Harold Kincaid

**Abstract** After opening with some metaphilosophical preliminaries sketching a naturalist framework that guides the paper, I devote my discussion to identifying dead and live issues in the traditional individualism-holism debate. The third section discusses standard reductionist theses about theory reduction. Arguments given for and against such claims has been conceptual in nature and thus to my mind misguided. However, the empirical evidence against reducibility now seems overwhelming. More dead ends will be the topic of the fourth section, where I will discuss claims that society does not exist and claims that social mechanisms require accounts in terms of individuals. In the last Section I look at numerous places in the social sciences where there are interesting open issues around the individualism-holism controversy. Those issues are about how holist or individualist we must or can be in senses I specify. The live issues in the individualism-holism debate are not global ones to be decided on general conceptual grounds but local and contextual empirical debates about how far we can get by proceeding without institutional and social detail.

### 8.1 Introduction

This paper is about what I take to be live and dead issues in debates over methodological individualism. Given that philosophers are unlikely to reach uniform agreement about what issues are dead or alive, perhaps this paper would be better described as about topics I find interesting vs. uninteresting. Of course, I am not just going to report my preferences—this paper will give arguments about why

---

H. Kincaid (✉)

School of Economics, University of Cape Town, Cape Town, South Africa  
e-mail: [harold.kincaid@uct.ac.za](mailto:harold.kincaid@uct.ac.za)

certain traditional individualism/holism claims are dead, or at least, not worth pursuing.

The paper will be divided into four subsections. The opening section will discuss some metaphilosophical preliminaries about how to approach the issues. The second section will discuss two standard reductionist theses, bad arguments for and against them, and explain why they are not worth pursuing. More dead ends will be the topic of third section, where I will discuss claims that society does not exist and claims that social mechanisms require accounts in terms of individuals. In the last, longest Section I look at numerous places in the social sciences where there are interesting open issues. Those issues, I will argue, are about how holist or individualist we must or can be. The live issues in the individualism-holism debate are not global ones to be decided on general conceptual grounds but local and contextual empirical debates about how far we can get by proceeding without institutional and social detail.

## 8.2 A Naturalist Approach

I begin with some philosophical preliminaries. My philosophical commitments are roughly naturalist in nature. For me that means that philosophy of social science is not independent of or prior to social research itself—the two interact and form a continuum. Social scientists discuss issues all the time that might be called philosophical or conceptual, and philosophers of social science take up topics that are quite empirical. Traditional projects of conceptual analysis looking for necessary and sufficient conditions tested against intuitions will not tell us much about the sciences we want to understand. Philosophy does not first decide what is ontologically real or essential and then the science commences. What is real and fundamental is itself a scientific question.

So in terms of the individualism/holism debate, I would argue that:

1. What the social sciences can or cannot explain is not decidable on purely philosophical grounds
2. Broad conceptual facts about social reality or about individuals are unlikely to get us far in the individualism/holism issue
3. Ontological issues concerning social entities are not first decided and then the social science is done accordingly

To flesh out this perspective a bit more, let me describe a positive alternative. When pursuing the individualism-holism debate we need to look at specific pieces of social research where the individualism/holism issues arise. Asking *in general* if social entities exist or if we can or cannot explain everything in terms of individuals is not helpful. Rather we need to identify concrete specific social science and behavioral science accounts and research and then ask about them what kind of entities we are committed to and to what extent those explanations are essential and sufficient. This means showing what individualism/holism debate would come to

case by case. Of course conceptual issues will arise in the process, but they must be disciplined by empirical specifics and originate from the empirical differences they make.

### 8.3 Classical Reductionism

With these preliminaries out of the way, let me turn to two traditional methodological individualist theses, namely, that:

- All social science explanations *must* be entirely in terms of individuals
- All social science explanations *can* be entirely in terms of individuals

The second claim about what can be done is entailed by the first about what must be done. If social science explanations have to be about individuals and we think social science is possible—and the vast majority of individualists do—then we are committed to the weaker thesis that we can explain in terms of individuals. So the rationales for the two theses will have to be interrelated but will not be identical.

These claim raise a number of points that ideally need to be clarified.

First, what are the restrictions on referring to individuals? Must it be in solely psychological terms? What kind of psychology? Standardly, individualists talk about belief, desires, attitudes, and the like. But behavioral science has gone much beyond these folk psychological models. What kind of cognitive science do we need to explain in terms of individuals? A related issue concerns how we characterize the relations between individuals. Only a radical individualist would claim that we can explain social phenomena entirely in terms of individual traits aggregated together. But how do we characterize the relations among individuals? Is a reference to individuals in terms of their role or position in institutions an individualist explanation?

A second aspect of these claims needing attention concerns what we mean by explanation. What is the relevant sense of explain? Is it a nomological -deductive notion? Some sort of “understanding”? A causal account?

The third element needing explication concerns what sense of “can” do we have in mind? If it is only a weak in principle claim about what is possible at the limit of scientific investigation sometime in the future—e.g. God could do it—then the claim has little relevance to current social science. On other hand, if it requires that we already have such explanations on the books, that is probably asking too much. Here I would go back to my naturalist perspective and think that we need to look at specific social explanations and possible individualist accounts of them to see how far they can go.

A final fourth concern is what motivates the idea that we must have individualist explanations? Must for what purposes? What is the rationale for saying how social science must proceed? I will return to this issue later.

I want to focus on the classic reductionist versions of the can thesis, viz that

It is possible now or will be in the near future to derive, at least approximately, any well-confirmed social generalizations or laws from accounts referring only to individuals and their relations

I think this thesis in this general form is no longer of much interest for two reasons: the standard arguments for and against it are of the broad conceptual kind that I labeled inadequate earlier and the relevant empirical evidence goes strongly against the thesis.

A standard argument for this version of individualism might be called the argument from composition and it goes like this:

If entities of kind S are composed of entities of kind I and their behavior is determined by the behavior of entities of kind I, then it follows that it is possible to explain what we know about S's in terms of I's

This argument or inference is very common—it is found in earlier writers like Watkins (1973) and by various recent advocates of “analytic sociology” with its emphasis on mechanisms (2009). The gut intuition is that since social institutions are just collections of individuals and not independent actors, then obviously they can be explained in terms of those individuals. The conclusion from the intuition is that explanations in social terms can be captured by explanations in purely individual terms and, assuming that explanations rely on generalizations of some sort, that the generalizations provided in social terms can be derived from generalizations about individuals. Such derivations produce standard reductions.

This inference wants to draw epistemic conclusions about what we must be able to explain from certain putative ontological facts about the world. However, our ability to explain depends on our conceptual resources and what they can do. Wholes are no doubt made up of parts, but that does not guarantee that our current understanding of parts allows us to capture our extant theories of wholes.

Actually the evidence from various sciences is that we generally cannot draw conclusions about the explanation of complex entities from information about their parts. Here are some areas where the reductionist ideal seems to have met its match:

*Quantum chemistry:* It is sometimes said that chemistry has been reduced to the fundamental physics of particles. However, quantum mechanical explanations of molecular structure only handle the simplest of cases. In general the chemistry of molecules has to be used and explanation in terms of orbitals, valence, bonding, etc. are needed for complex molecules that cannot be captured by quantum mechanics.

*Molecular biology:* It is likewise often thought that biology has largely been reduced to chemistry. Two key areas where it is often claimed that we have complete explanations in molecular terms are genetics and the functioning of the immune system. Yet attempts to identify genes with sequences of DNA have failed—Lenny Moss's *What Genes Can't Do* (2002) argues this decisively—because genetic inheritance is far more complicated than the optimistic beads on a string picture spawned by Watson and Crick. Inheritance is influenced, for example, by imprinting of histones which provides information over and above that



contained DNA sequences. But population genetics and other parts of biology get on quite well with the traditional notion of a gene. In the case of the immune system antibodies are essential elements in explaining immunity, and we have molecular explanations of how they function. However, the chemical structure of antibodies is amazingly diverse, and in the end to be an antibody is to activate the immune system. So we cannot explain everything in terms of the antibody molecular structure—appeal to the characteristics of the immune system are essential.

*Complex dynamic systems:* On a much more general note, studies of complex dynamic systems repeatedly show that macro patterns emerge from individual dynamics in ways that cannot be predicted from the nature of individual agents. Of course, this reflects a limitation of our knowledge, but reductionist claims presuppose that our knowledge can suffice to explain.

So the compositional argument for classic reducibility fails. I now want to argue that some standard arguments for *irreducibility* have similar problems. A common argument on the holist side is that explanations in terms of individuals in the social sciences must explain in terms of social roles and therefore presuppose explanations in terms of social entities. To use an example from my own work, “there are no prisoners without prisons” (1996). Roles allegedly cause problems for reductions because roles are only defined against a backdrop of institutional and organizational structure.<sup>1</sup> So while the role is occupied by an individual, the role descriptors are about more than individuals, namely, the social entities in which they exist. Organizational and institutional structures are social processes or entities that individualist should explain, not take as given, if reduction is to succeed.

Zahle (2007) argues convincingly that these kinds of arguments are inclusive or question begging. In my own defense I think I flirted with such arguments but never consummated the relation as it were. The more defensible view is that we must look case by case at alleged explanations of social phenomena in terms of individuals to determine to what extent the explanations avoid reference to institutions, etc. Not all roles and invocations of them have the same social import. So as an all purpose defense of holism, such arguments overreach. As a local problem they may be compelling. Thus I do not rule out that there might be local reductions—reductions of quite specific social accounts to individualist accounts—that are informative and valuable.

---

<sup>1</sup> I refer to both organizations and institutions because these terms are used differently. Some think of institutions as just norms and norms as just regularities in behavior. I have in mind a richer (and I think more realistic) notion of an institution.

## 8.4 More Dead Ends: Mechanisms, Eliminativism, and Explanation

I want next to look at three more theses in the individualism-holism debate that on my view are not worthy of further debate: the claim that we must have individualist *mechanisms*, the idea that purely social explanations are incomplete and the view that social entities are fictions. I find these debates dead ends because they are either undecidable at this level of generality or they are too easily decidable to raise interesting questions about the social world.

Elster (1983) has claimed that we need individualist mechanisms in the social sciences to rule out confounding causes. There has been much rhetoric from social scientists and philosophers about needing mechanisms, echoing Elster's original claim.

Most of this literature has been quite unclear about what mechanisms are and why they are needed. I am suspicious of any blanket claim about mechanisms in the social sciences for two reasons. First, I am suspicious of broad methodological pronouncements in science in general. In practice methodological rules require domain and context specific knowledge for their interpretation and application (Day and Kincaid 1994; Kincaid 2012b). Simplicity, for example, has a role in science, but the work that it does often comes from domain-specific instantiations that embody substantial empirical claims (Sober 1989). I would expect the same for claims about mechanisms in the social sciences. A second reason for skepticism about the demand for mechanism results from the fact that the claim can be given several different readings and motivations which are logically independent and need not stand or fall together.

A framework for thinking about those differences would help clarify the issues, and I turn now to sketch out the logical space of claims (Kincaid 2012a). A first question is what we want mechanisms for. As I argued some time ago (1996, 1997), we might want a mechanism for *explanatory* purposes or for providing *evidence*. These need not be the same. I may have a well-confirmed association or even a causal claim, but think it is not sufficiently deep enough to explain in some sense—I may think that to explain I need to know how the relation obtains. On the other hand, I might want mechanisms because I doubt an association is real in the first place and believe that providing a mechanism would lend it further credibility. So explaining and confirming with mechanisms can come apart.

We can also use mechanisms to confirm two different types of causal claims: assertions that a causal relation *exists* vs assertions about the *size* of the relationship. It is one thing to know that C causes E, another to know how changes in the various values of C result in differing values of E. Mechanisms might be valuable for determining effect size but not effect or vice versa. So having a mechanism might increase my evidence that changes in the interest rate cause changes in employment or it may be needed for me to infer how much a change in interest rates increases or decreases employment.

Another important difference in thinking about mechanism turns on whether we want *horizontal* or *vertical* mechanisms. Asking for horizontal mechanisms is asking for the steps that led from the cause to the effect—the intervening causes that makes for a continuous process. I label these horizontal because they are mechanisms at the same level as what the mechanisms relate.

In contrast, vertical mechanisms are the underlying component parts that gives the cause its capacities to produce the effect. So if I explain how changes in the interest rate cause changes in employment by citing changes in aggregate investment, I am citing a horizontal mechanism. If I explain changes in aggregate demand that cause changes in price by citing the budget constraints and utility functions of individuals, then I am citing a vertical mechanism. Arguments for mechanisms in the social sciences are often unclear about which sort of mechanism they entail.

Orthogonal to all the above distinctions, we might think that having a mechanism is *essential* to what we want to do or we might more modestly claim that having them is helpful or *useful*. Is it the case that we cannot do without mechanisms or rather that they make things easier in scientific practice? Assertions that something is essential for good science make for much stronger demands than claims that they are of some benefit.

So the logical space of claims about mechanisms is large. As a result there are really many different things one might be claiming in asking for mechanisms and surely their plausibility has to be decided case by case and probably by domain and discipline as well. To evaluate demands for mechanisms in the social sciences we need to get specific about exactly which claim we are making among the many possibilities.

Elster's claim then is that we need vertical mechanisms—the details realizing social entities—in terms of individuals to confirm. But it is obvious that this cannot be right as a general principle. If I throw a baseball into a glass window and it breaks, then if nothing else hit the window, I have good evidence that the moving ball broke the window. I don't need the molecular details of the ball or the glass to confirm this. So the general principle is mistaken.

That leaves us with the question whether we have parallels to the baseball in the social world. I think we do. Here is an example I have relied upon before. Two very impressive sociologists, Hannan and Freeman (1989), did a study of the differential survival of organizations according to environments they faced. They showed that different kinds of organizations—high tech firms, restaurants—survive better in different kinds of social environments. Some have “generalist” strategies in the face of variable or unstable environments. Others have specialist strategies in different environment. The work of Hannan and Freeman is quite rigorous. Here we have social entities—institutions and organizations—interacting with their social environment. Hannan and Freeman do not give us individual level detail, but they do control for other social variables such as the existence of other kinds of institutions, government support, and the like. Like balls striking windows, we don't need the underlying details to have reasonable evidence.

A further difficulty with the claim that we must have mechanisms in terms of individuals is that the social sciences may want to provide vertical mechanisms, but

not mechanisms in terms of *individuals*. Macroeconomic explanations, for example, might have realizing mechanisms but the mechanisms might be in terms of the behavior of firms, for example.

I should note that social science research like Hannan and Freeman's provides compelling evidence against the individualist thesis that explanation *must* be in terms of individuals. Describing causes of phenomena is surely one solid sense of explaining. Hannan and Freeman describe the causes of different survival of organizations and take the usual steps to rule out confounders. To this extent they have explained without invoking the behavior of individuals.

On the other hand, I think no reasonable holist will deny that knowing individual detail can provide for fuller explanations—the thesis that social explanations are in a sense incomplete without individual detail is quite plausible. Thus this is another thesis that is no longer a live issue because the evidence is overwhelmingly in favor of it. What makes for a fuller explanation or measures its incompleteness? One way to flesh this out is terms of the number of questions answered. Another is in terms of the depth or denseness of the causes cited. Our baseball breaking the window example is a case in point: we know the causal relation but the molecular details of the hardness of the ball and the brittleness of the glass are nontrivial issues in physics and chemistry. Understanding those would of course increase the number of questions we could answer and provide us with more causal detail. Providing individual level detail in social explanations surely does the same.

So to make my position on individualist mechanisms clear: there is no general demand for mechanisms that is justifiable. Once again at best there are local cases where individualist mechanisms in one or another of the different senses of mechanism can have real virtues.

A third dead or at least uninteresting issue concerns a kind of eliminativism about social entities. It's most famous representative is Maggie Thatcher. Her view is that social entities do not exist. Actually her view is a bit more complicated—she says “there is no such thing as society. There are individual men and women, and there are families” (1987). No reasonable holist is going to claim that social entities exist over above individuals in complex relations. Similarly no reasonable individualist is going to deny that there are complex aggregations of individuals. I think the core issue is not whether social entities exist—they do—or whether they exist independently of individuals—they do not—but really is still about what we need to use in our explanations and what counts as evidence for them.

But I grant that there are unreasonable holists as well as unreasonable individualists. This quotation from Ron Sun (2005) in a recent book on collective intelligence is a nice example of an unreasonable holist whose claims would understandably incite individualist inclinations:

regarded as an emergent property, collective cognition is *holistic* in the sense of being essentially macroscopic rather than a mere summation of microscopic local properties. A group belief is something that transcends the sum of the individual members' beliefs (5232 Kindle).

The problem with such claims is not ontological sins on my view. I would take a Dennett style intentional stance approach and ask to what extent we can pick out real patterns by invoking group beliefs. If we treat groups as having beliefs and as subject to rationality constraints, can we predict their behavior in the relevant environments? If so, we have identified something real. The problem with Sun's appeal to group beliefs, however, is a complete lack of explanatory integrity: how do I measure group beliefs? What concrete explanatory role do they play—can we show they have some identifiable causal influence or are influenced by other identifiable factors? Can we say how they relate to individual beliefs? Sun does almost nothing to answer these questions.

## 8.5 How Holist or Individualist Can or Must We Be?

I want to turn now me from beating dead horses to much more interesting pursuits. There are numerous ongoing social science controversies that can reasonably be seen as instances of an individual-holism debate. Most of these can be phrased as debates about how holist or individualist can or must we be? This question is roughly about how much social structure we need to add to facts about individuals in order to successfully explain social phenomena? I take “social structure” to run the gamut from relatively thin social roles—a is the neighbor to b—to the full fledged invoking of large scale social entities, e.g. nation states maximizing their interests in international relations with other nation states, with lots of mixes in between these two extremes. I think there are multiple cases where the question how holist or individualist we must be are core empirical issues in the social sciences. I discuss some examples in the rest of the paper.

Of course, the standard individualist answer to my question is that we can be entirely individualist—we can completely eliminate or reduce explanations in terms of social entities to accounts in terms of individuals. The standard holist answer is the corresponding denial that we can ever be so individualist. I think the evidence is quite telling against the blanket individualist claim but that the holist alternative is implausible as well. Still even where it may be clear that we cannot be completely individualist or holist, the interesting question remains about just how much social explanation do we need. I give a number of examples of such debates in what follows.

Rational choice game theory is certainly in the individualist spirit. Social outcomes are explained in terms of individuals pursuing their best outcomes, given what everyone else is doing. However, even at its most individualist, it presupposes some social structure to get going. While it explains in terms of individuals interacting, it has to take as given individual preferences, knowledge about the possible strategies of others and their payoffs; it leaves those unexplained. These are things that have natural social explanations in terms of socialization, institutional rewards and punishments, etc. In that sense rational choice game theory is only somewhat individualist. However, recent developments suggest that even

more social structure is needed. Many realistic games that individuals must play turn out to have multiple equilibria, given the constraints assumed on preferences, payoffs, and the like. How are such multiple possibilities reduced to one unique outcome in reality? For many games there seem to be focal points—socially salient choices—that allow people to coordinate. These focal points come from social norms and institutional arrangements that are not derived from the rational choice game theory models. So in this sense rational choice game theory must on occasion be even more “holist”.

Agent-based and evolutionary game theory models are alternatives to traditional rational choice accounts and they also raise the question of how individualist we can be. Evolutionary game theory does not suppose that individuals are rationally maximizing preferences but only that individuals have different behavioral strategies and that those strategies interact in ways that lead to equilibrium outcomes in populations. Evolutionary game theory generally assumes that individuals interact randomly. Agent based or network models invoke a similar framework but add in varying levels of constraints on how individuals interact. The evidence seems to be that some phenomena can be modeled with minimal constraints and that others require significant appeal to social structure.

For example, Alexander in his book *The Structural Evolution of Morality* (2010) looks at cultural evolutionary accounts of the social phenomena. He shows that getting trust and cooperation as stable outcomes requires models with significant social structure. Here social structure means restrictions on who interacts with whom. Natural social examples include categories such as race, ethnic identity, social position, etc. So here we need to be somewhat holist—how much so depends on the specific phenomena we are looking at.

An early example of agent based modeling that is in the individualist spirit is Schelling’s classic model of residential segregation. It works from individuals interacting based on their preferences and shows that racial segregation could result merely from individual preferences. However, extensive social research shows that institutional factors such as redlining by banks are probably an important part of the story. Again it is an empirical issue how holist or individualist we must be.

Another interesting example from agent based modeling comes from Padgett and Ansell (1993) in a classical article. They apply network models to the society of the Medici family. They show that in one sense social structure played a minimal role—it is ties between families, not the wealth of families, that explain the Medici’s influence compared to those they were competing with. However in other ways we have to be more holist: crucial to their network account is the existence of an elite—so network ties are within that group—and the influence of wars, more sociological level phenomena.

Another area where a key question is how holist or individualist must we be comes from studies of personality and character—how much does individual personality explain behavior as opposed to incentives, norms, sanctions and the like? Doris in his book *Lack of Character* (2002) provides a convincing survey showing that often character is a weak explanatory variable compared to institutional incentives, expectations, etc. There is an older sociology literature pointing to

similar conclusions that argues attitudes are weak predictors of behavior. However, character and attitudes are concepts with deep folk psychological origins. Contemporary psychometric notions of personality may well have more predictive power.<sup>2</sup>

Yet, as I been keen to point out, these kind of issues have to be taken case by case. Mallon and Kelly (2012) survey the attempts to explain race as a social role. It has been commonplace now to say that race is biologically meaningless and is only a social construct—the result of institutional racism. However, they argue plausibly for a more moderate middle position. Race does not exist as a genetic phenomena in that our many different racial categories—what counts as black in Brazil, South Africa, and the US are rather different things—do not tie closely to genetic differences. But Mallon and Kelly argue that nonetheless there is more to race than social categorization. There is extensive social psychological and even neurobiological evidence showing that ingroup favoritism and outgroup biases come quite naturally to us. Some of the evidence is from social psychological experiments—tell people they are in a group and their judgments of those who are not in the group are biased. Put people in a brain scanner whose attitudes and behavior is nonracist and show them imagines of blacks and whites. You see fear regions in the brain much more active when the black faces are present than the white ones.

This is clearly a case where it is useful to be more individualist. I said earlier that a trivial individualist claim was that individualist details were needed for full explanation. However, even that pronouncement was probably too general. In the racism case the need for individual factors to fully explain is quite strong. However, elsewhere the requirement may be less demanding. Gary Becker showed a long time ago that the well- established phenomenon of the aggregate demand curve slopping downwards did not have to be explained in terms of individuals maximizing their consumption possibilities. Random choices would produce the same results, given budget constraints. In this case the individual details that are needed are quite thin, or at least much less thinner than that assumed by traditional economic models of consumption. Becker derived his results analytically, but there is also experimental evidence that other economic phenomena depend on institutional details rather than the rationality of agents. Gode and Sunder (1993) ran experiments with dumb agents (robots) placing bids in continuous bid/ask double auction. Their agents found the competitive equilibrium—standard rationality assumptions are unnecessary because of the institutional features of the auction.

Issues concerning how strong our assumptions about individuals must be to explain social phenomena are central to current ongoing debates about the sources of social cooperation and altruism. Is cooperation explained by other regarding preferences of individuals and willingness to undertake costly punishment or are thinner, more standard self interested motives sufficient when combined with repeated social interactions that enduring institutions represent? These are

---

<sup>2</sup>Thanks to Don Ross for noting the difference between folk psychological notions of character and more scientific notions of personality.

fundamental questions about the nature of human social and cooperation and they turn on the question of how individualist must we be.<sup>3</sup>

The question how individualist we can be takes on further interesting empirical focus when it comes to explaining irrational behavior with social ramifications. Here individualism as a methodological imperative or reductionist claim runs into the problem that individuals may not best be explained *as individuals*. Behaviors such as addiction, extreme time discounting, and risk aversion would seem to be important details to add to social explanations of related phenomena. Yet there are serious arguments that such behaviors need to be explained in terms of subpersonal agents—we need to postulate subselves to explain the behavior of whole human beings. One line of this thinking is called piceconomics and its best known representative is the psychiatrist George Ainslie. I want to make no guesses about how well this research program is going, but I mention it to stress my point that explanation “in terms of individuals” is not a univocal goal. Which level of lower level detail is needed is an empirical issue and not one that has a uniform, universal answer.

I have been putting the question of how individualist or holist must we be in terms of whether we need social variables or factors in addition to individualist ones. However, a further gloss on this question would be not to just ask if social factors have a causal role but how large that role might be. In other words, we can explain by showing that some factor has a causal effect, but we can add to that explanation by citing the size of that effect.

The question of how large a causal role social factors play does at least indirectly come up in some of the work cited above, but asking it presupposes some common metric for comparing importance. Some recent work in the methodology of causal inference suggests how that common metric can be (and is being) constructed in the social sciences. Multilevel structural equation models are the tools I am thinking of. Structural equation models can be used to explicitly define a set of causal relations and then test the dependencies and independencies they imply against the data. Multilevel structural equation models describe variables at different level of aggregation or grain, e.g. they relate student outcomes to schools, school districts, etc., and they can posit explicit causal relations between variables at different scales and as well as among variables at the same scale. Such models can estimate the relative causal influence of the factors in the model.

Multilevel structural equation models raise a last issue that I want to finish with that goes beyond the question about how holist or individualist can we be, viz., how do accounts in terms of individuals and social entities interrelate? The best literature in the mechanisms literature—and I argued something similar a long time ago about molecular biology—is that there needs to be integrated interlevel accounts. Philosophers of biology have made some progress on what that looks like in parts of biology. But the story is still a very much in need of detail in the social sciences I think. Multilevel structural equation models assume we can describe how causal

---

<sup>3</sup> See Guala (2012) for some of the issues.



processes interrelated. But it is often quite unclear that we can do that. How would Hannan and Freeman's macrosociological account in terms of differential organizational survival and competition relate to game theory or agent based models of institutions? How does the classical theory of the firm as a maximizing agent relation to individual or suborganizational accounts in terms of incentives and the like?

It seems to me that this issue about integrating the individual and the social is a very live one across many areas of the social sciences. So while my initial skeptical statements at the beginning of the paper about the limits of philosophy and philosophy of social science may have seemed liked calls for a mass layoffs of philosophers, my conclusion is far different. There is enormous philosophical and conceptual work to do, but it has to be in employment directed towards making better social science.

**Acknowledgments** Finn Colin, Julie Zahle, and Don Ross made careful comments that much improved this paper.

## References

- Alexander, J. (2010). *The structural evolution of morality*. Cambridge: Cambridge University Press.
- Day, T., & Kincaid, H. (1994). Putting inference to the best explanation in its place. *Synthese*, 98, 271–295.
- Doris, J. (2002). *Lack of character*. Cambridge: Cambridge University Press.
- Elster, J. (1983). *Explaining technical change*. Cambridge: Cambridge University Press.
- Gode, D., & Sunder, S. (1993). Allocative efficiency of markets with zero intelligence traders. *Journal of Political Economy*, 101, 119–137.
- Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, 35, 1–15.
- Hannan, M., & Freeman, J. (1989). *Organizational ecology*. Cambridge: Harvard University Press.
- Hedstrom, P., & Bearman, P. (2009). *The Oxford handbook of analytical sociology*. Oxford: Oxford University Press.
- Kincaid, H. (1996). *Philosophical foundations of the social sciences: Analyzing controversies in social research*. Cambridge: Cambridge University Press.
- Kincaid, H. (1997). *Individualism and the unity of science: Essays on reduction, explanation, and the special sciences*. Lanham: Rowman & Littlefield.
- Kincaid, H. (2012a). *Oxford handbook of the philosophy of the social sciences*. Oxford: Oxford University Press.
- Kincaid, H. (2012b). Naturalism and the nature of evidence in economics. In U. Mäki (Ed.), *Handbook for the philosophy of science: Philosophy of economics* (pp. 115–134). Amsterdam: Elsevier.
- Mallon, R., & Kelly, D. (2012). Making race out of nothing: Psychologically constrained social roles. In H. Kincaid (Ed.), *Oxford handbook of the philosophy of social science* (pp. 507–533). New York: Oxford University Press.
- Moss, L. (2002). *What genes can't do*. Cambridge: MIT Press.
- Padgett, J., & Ansell, C. (1993). Robust action and the rise of the Medici, 1400–1434. *The American Journal of Sociology*, 98, 1259–1319.
- Sober, E. (1989). *Reconstructing the past*. Cambridge: MIT Press.

- Sun, R. (2005). *Cognition and multi-agent interaction: From cognitive modeling to social simulation*. Cambridge: Cambridge University Press.
- Thatcher, M. (1987, October 31). *Woman's own*, 8–10.
- Watkins, J. (1973). Methodological individualism: A reply. In J. O'Neill (Ed.), *Modes of collectivism and individualism* (pp. 179–185). London: Heinemann.
- Zahle, J. (2007). Holism and supervenience. In S. Turner & M. Risjord (Eds.), *Handbook of philosophy of science: Philosophy of anthropology and sociology* (pp. 311–343). Amsterdam: Elsevier.

## Chapter 9

# Explanatory Strategies Beyond the Individualism/Holism Debate

Jeroen Van Bouwel

**Abstract** Starting from the plurality of explanatory strategies in the actual practice of social scientists, I introduce a framework for explanatory pluralism – a normative endorsement of the plurality of forms and levels of explanation used by social scientists. Equipped with this framework, central issues in the individualism/holism debate are revisited, namely emergence, reduction and the idea of microfoundations. Discussing these issues, we notice that in recent contributions the focus has been shifting towards relationism, pluralism and interaction, away from dichotomous individualism/holism thinking and a winner-takes-all approach. Then, the challenge of the debate is no longer to develop the ultimate individualistic approach or defending the holist approach, but rather how to be combine individualism and holism; how can they co-exist, interact, be integrated or develop some division of labour, while making the best out of the strengths and limitations of the respective explanatory strategies of holists and individualists? Thus, the debate shifts to how exactly pluralism should be understood as the next leading question, going beyond the current individualism/holism debate. The paper ends with a discussion and evaluation of different understandings of explanatory pluralism defended in the literature.

---

J. Van Bouwel (✉)  
Ghent University, Ghent, Belgium  
e-mail: [Jeroen.VanBouwel@UGent.be](mailto:Jeroen.VanBouwel@UGent.be)

## 9.1 Introduction: The *Topics, Methods and Aims* of the Individualism/Holism Debate

Central in the individualism/holism debate figures the idea of *methodological individualism*. It “amounts to the claim that social phenomena must be explained by showing how they result from individual actions” (according to the *Stanford Encyclopedia of Philosophy*). Its contender, *methodological holism*, negates this claim and defends that good social explanations may also invoke social structure, culture or social functions without references to individual actions. We all know there exist many variations of methodological individualism and holism, but I will not spell them out here (see, e.g., Udehn 2001). Rather, I will focus on some of the characteristics of the individualism/holism *debate*.

It is striking how often scholars commenting on the debate seem to be dissatisfied about it; calling it, e.g., a “notoriously unfruitful controversy” (Ylikoski 2012: 21), “confused” (Zahle 2006: 312), or associating it with “despair” and “frustration” (Bhargava 1992: 5). Whatever the exact reasons for this dissatisfaction, I do think the debate would benefit greatly from making the *topics, methods* and *aims* of the debate more explicit (and giving up the *winner-takes-all* approach, discussed in Sect. 9.2 below). While participants must not agree unanimously what the topic, method and aims are or should be, each participant could at least be explicit about her specific angle to the debate.

Considering past contributions to the individualism/holism debate, different options have been explored:

- *Topics*: Is the debate focussing on methodological individualism, ontological, political, logical, semantic, legal, epistemological, or axiological individualism, . . . ; contrasted with holism or collectivism – methodological, ontological, or political, . . . (see, e.g., Bunge 2000)?
- *Methods*: On which basis do we analyse the topics? On the basis of intuitions, metaphysical commonplaces, conceptual analysis, logical discussion, transcendental arguments, political convictions, importing philosophy of mind machinery, or analysing social scientific practice, . . . ?
- *Aims (and scope)*: Should the debate lead to elaborating one’s own social theory, to a social ontology to be adopted by all social scientists, to elucidating scientific practice, or to improving scientific practice, . . . within a specific approach, or a discipline, or the social sciences at large?

Clearly explicating which *topic(s)* one is discussing, might avoid mixing up *topics* as frequently happens in the debate: arguments for *political* individualism sometimes automatically imply a defence of *methodological* individualism; *ontological* holism is deemed to immediately follow from advocating forms of *methodological* holism; and, ontological individualist arguments are used to prove methodological individualism right (cf. Sect. 9.3 below). Further, the lack of clarity on what individualism or supra-individualism, holism and collectivism exactly means level-wise, i.e. on what ‘level’ it can be found and which other ‘levels’ are

in play, is also a source of confusion (for holism versus collectivism, see e.g. Pettit 1993). Thus, clarifying the topic(s) seems imperative.

Next, one should decide how to investigate and argue for, i.e. what *method* to use in scrutinizing the topic(s). Method has not received enough attention in the debate, although the recent popularity of the field of social ontology does trigger questions about method. Kincaid (2012), for instance, wonders whether philosophers of science, scrutinizing the practice of the scientists (e.g. biologists), would ask them to follow the philosopher's ontological *a priori* speculations, just like some philosophers of social science seem to do. Kincaid questions methods like the one advocated by Searle that give primacy to ontological reflections on the basis of conceptual analysis developing "a clear conception of the nature of the phenomena" before turning to methodology and social research practices, cf. "social ontology is prior to methodology" (Searle 2009: 9).

Finally, why are we having these debates? What are the *aims* and the *scope*? These questions are hardly ever addressed in the debate. One has the impression that some are more concerned with developing (and defending) their own social theory, while others want to describe the actual assumptions of social scientists (or of a particular approach within social science) vis-à-vis the individualism/holism debate. A third group aims at stipulating normative guidelines for a better social science or improving a particular approach or social theory. There might be other aims as well, what I want to point at here is that the aims of the debate deserve more explicit discussion in order to improve the debate's sharpness and focus.

In this paper, I take the following position to discuss the individual/holism question:

- *Topics*: Here, I focus on methodological/explanatory individualism and holism (with nuances about the 'levels' of explanation).
- *Method*: I start from the actual explanatory practice of social scientists, which I aim to clarify and evaluate on the basis of philosophy of science literature.
- *Aims and scope*: We should not have the debate primarily for philosophy's sake, but for a social science able to serve multiple aims adequately (cf. the different explanatory interests of scientists) and I hope my approach is useful for social scientists themselves.

Starting from this position, I highlight the plurality of explanatory strategies (on several levels) in scientific practice and advocate explanatory pluralism (as a normative endorsement or legitimization of this plurality) in Sect. 9.2. Based on the insights of this section, I revisit some key issues in the individualism/holism debate, namely emergence, reduction and the idea of microfoundations in Sect. 9.3. Then, these key issues are integrated in a discussion about the different understandings of explanatory pluralism in Sect. 9.4. The exact understanding of pluralism is introduced as the next leading question, going beyond the current debate. Section 9.5 concludes.

## 9.2 Explanatory Strategies in Social Scientific Practice

### 9.2.1 *Introducing the Framework for Dealing with Plurality in Explanatory Practice*

Across the social sciences we find a plurality of ways in which social scientists try to explain social phenomena. In order to deal with this plurality, I have developed a framework for understanding explanatory plurality in scientific practice.<sup>1</sup> The framework works as a tool to (a) *make the explananda as explicit as possible*, and (b) pay attention to the *underlying explanatory, epistemic interests*. This is imperative for clarifying discussions about competing explanations: there are many cases where two explanations of the same phenomenon are perceived as competitors, but actually have different *explananda*. The framework employs the erotetic model of explanation that regards explanations as answers to why-questions.<sup>2</sup> Making the explananda as explicit as possible as well as paying attention to the different epistemic interests, can be done by explicating the explanation-seeking questions.

Analyzing social scientific practice, different explanation-seeking questions or requests can be distinguished. I do not consider the questions and motivations mentioned here as the only possible ones, but I do believe they are omnipresent in social science practice. At least five types of explanatory questions can be distinguished:

- (E) Why does  $x$  have property  $P$ , rather than the expected property  $P'$ ?
- (I) Why does  $x$  have property  $P$ , rather than the ideal property  $P'$ ?
- (I') Why does  $x$  have property  $P$ , while  $y$  has the ideal property  $P'$ ?
- (F) Is the fact that  $x$  has property  $P$  the predictable consequence of some other events?
- (H) Is the fact that  $x$  has property  $P$  caused by a familiar pattern or causal mechanism?

First, explanation-seeking questions can require the explanation of a contrast, e.g. of the form (E), (I) and (I'). Contrastive (E)-type questions, for instance, can be motivated by surprise: things are otherwise than we expected them to be and we want to know where our reasoning process failed (which causal factors did we overlook?). Contrastive questions of type (I) and (I') can be motivated by a therapeutic or preventive need; they request that we isolate causes which help us to reach an ideal state that is not realised now, comparing the actual fact with the

---

<sup>1</sup> This framework was developed by analyzing discussions about 'the best explanation' among social scientists with case-studies done in sociology, economics, international relations, history, medical science, etc, see, e.g., Van Bouwel (2003, 2004b), Van Bouwel and Weber (2002a, 2008a), Weber and Van Bouwel (2002).

<sup>2</sup> For more details on the erotetic model, see, e.g., Garfinkel (1981), Kincaid (1997), Risjord (2000) and van Fraassen (1980).

one we would like to be the case (therapeutic need) or to prevent the occurrence of similar events in the future (preventive need).

The form of a *contrastive* explanation (i.e., an answer to a contrastive question) enables us to obtain information about the features that differentiate the actual causal history from its (un)actualized alternative, by isolating the causes that make the difference; this information does not include information that would also have applied to the causal histories of alternative facts.

Second, non-contrastive explanation-seeking questions, concerning plain facts, like (F) and (H), are also omnipresent in social science. These non-contrastive questions can have different motivations. One possible motivation is sheer intellectual curiosity, with a desire to know how the fact “fits into the causal structure of the world” or to know how the fact was produced from given antecedents via spatio-temporally continuous processes. A more pragmatic motivation is the desire to have information that enables us to predict whether and in which circumstances similar events will occur in the future (or the anticipation of actions of persons/groups). Another possible motivation concerns causally connecting object  $x$  having property  $P$  to events we are more familiar with.

The form these explanations of *plain facts* (answers to non-contrastive questions) have, shows how the observed fact was actually caused, which implies providing the detailed mediating mechanisms in a (non-interrupted) causal chain across time, ending with the explanandum, or –considering the second motivation – the explanation can follow a covering law/law-based model.

By making the different possible explanation-seeking questions explicit, the motivation – explanatory interest – and the explanatory information required will be taken into account. Given that one social phenomenon can be the subject of different questions, and that we want to answer these different kinds of explanatory questions in *the best possible way*, different forms of explanation are indispensable. In order to decide on *the best possible way*, we consider (a trade-off between) the criteria (a) *accuracy* – relation with reality, precise description, (b) *adequacy* – relation to what the explained expects from the explanation addressing the explanatory interest, and (c) *efficiency* – amount of work and/or information needed for the explanation. To clarify these criteria and the idea that there often is a trade-off between them, let us compare explanations with maps. A subway map like the one of the Paris Metro is *adequate* for its users because it *accurately* represents specific types of features (e.g. direct train connections between stations, number of stations between two given stations, ...) while other features are consciously *less accurately* represented (the exact distances between the stations, the relative geographical orientation of the stations, ...). If the latter would be represented more accurately, the map could become less *adequate* for its intended users and a perfectly accurate representation mirroring every detail would be utterly useless. Furthermore, one could make the map more accurate, less adequate (without being completely inadequate), but also a lot less *efficient* in use (e.g. by making it less abstract, providing more cumbersome, obsolete information or by being too demanding or complicated to use). Other maps (e.g. Paris’ shopping or tourist attractions maps) require other kinds of information (relating to, e.g., distances,

details about street names, house numbers, etc.) in order to be useful – the best trade-off between accuracy, adequacy and efficiency differs depending on the interests or desiderata at play. Thus, on the one hand, because of different interests or desiderata, it is impossible to make a map that is ideal in all possible situations. On the other hand, not all maps are equally good, as one can make claims of superiority that are bound to specific situations. The same can be said for forms of explanation.<sup>3</sup>

Summarizing, an explanation is an answer that should be evaluated in relation to a question that is a specific request for information (and the precise meaning of the question is therefore important). Making the explanation-seeking questions as explicit as possible may show that, given that explanatory interests and contexts select distinct objects of explanation, a (apparently) similar question about one social phenomenon, results in very different questions and answers in which the most *accurate*, *adequate* and *efficient* explanatory information (in relation to the explanatory interest) is provided. Hence, different forms of explanation on different levels are indispensable to answer the respective explanation-seeking questions in the best possible way.

As concerns the debate between methodological individualists and holists this implies, first, that the claims of methodological individualists are not tenable, and, second, that the claims of methodological holists should (at least) be qualified, specifying to what extent outspoken individualist and reductionist explanations are allowed. Methodological holists have been focussing mostly on formulating arguments against methodological individualism and they have not invested enough in developing their own ideas of what a satisfactory explanation looks like. Furthermore, when debating, individualists and holists have adopted similar yet flawed ways of reasoning, as I elaborate below. I hope to go beyond these ways of reasoning by introducing the framework and focusing on explanatory strategies. Let us first give an example of how the framework can be used in dealing with social scientific practice.

### 9.2.2 *The Framework in Social Scientific Practice*

Graham Allison's classic study *Essence of Decision* (1971; Allison and Zelikow 1999) offers an interesting example to briefly illustrate how the framework just introduced helps us to understand plurality. Allison provides us with three different models to explain the Cuban Missile Crisis (1962). These are:

- Model I: *Rational Actor Model*, in which unitary nation states act on a rational basis;

---

<sup>3</sup> Also see Van Bouwel and Weber (2008a, b) for more about these criteria.



- Model II: *Organizational Process Model*, which opens the black box of the unitary state and points at the myriad of organizations constituting the state (driven by the logic of organization instead of a logic of optimization/maximization);
- Model III: *Governmental Politics Model*, which zooms in on actual people that make up states and organizations, their personal power, networks, skills of persuasion, etc.

Allison himself does not offer many instructions on how to deal with the plurality of explanatory models. Is one of the three models the correct one, does one have to add them up to get a satisfactory explanation, do they cancel each other out? The framework introduced above provides us with a satisfactory solution for dealing with the plurality.<sup>4</sup> None of the three possible ways just mentioned to deal with the problem of the plurality of explanations convinces: picking one model as the best one is not desirable, because none of the models performs well for all possible explanatory interests; adding up the three models also fails, because this means using models to satisfy interests they are not suited for; discarding all models is not an option either, because the models do succeed in satisfying some of the explanatory interests. How to make sense, then, of the plurality of explanations? The solution is to systematically choose the model that best serves the epistemic interests as made explicit in the explanation-seeking question.

Model I, the *Rational Actor Model* is apt for answering the questions (F) *Is the fact that x has property P the predictable consequence of some other events?* and (H) *Is the fact that x has property P caused by a familiar pattern or causal mechanism?* As concerns (F)-questions, predictions demand a model that makes law like statements. For this, the statements must be general and necessary. From Model I it could be inferred for example that whenever there is a missile gap between countries and these countries have a disagreement, the weakest country will have a strong desire to close that gap. This statement is both general enough and gets its necessity from the underlying expected utility calculus which yields an unambiguous solution. Dealing with (H)-questions, the coarse-grained, unrealistic nature of Model I is compensated by its ability to bring any situation down to a simple calculus. In this model, the USSR wondering whether or not to put nuclear missiles in Cuba is in all Model I respects similar to being at a bakery pondering about whether to have just bread or to go for the croissant. As such, Model I is by far the best option for creating a sense of familiarity. Thus in answering the (F) and (H) questions, Model I will focus on the desire of the USSR to close the missile gap. As such, this desire creates familiarity and the level of analysis on which it is situated allows for regularities.

Model II, the *Organizational Processes Model* addresses questions of the form (E) *Why does x have property P, rather than the expected property P'?* very well. As the actions emerging from large organizations can take very strange, unfamiliar

---

<sup>4</sup> In an earlier paper, I extensively show how the framework can be used to deal with this question, see De Langhe et al. (2007).

forms due to organizational biases, and raise serious doubts concerning the rationality of the organizational process as a whole, unexpected events can be explained as outcomes of long and slow processes of organizational struggle, often resulting in actions nobody ever wanted; or the presence of ‘standard operating procedures’ (SOPs) which were designed not for the present situation but for some previous circumstance.

An example of Model II satisfying the (E)-interest is the following: *Why did the USSR decide to place offensive missiles in Cuba without camouflaging the nuclear sites during construction, while they did so (only) after U-2 flights pinpointed their locations?* The organizational processes model explains this unexpected aspect the best. The implementation of the USSR decision is assigned to organizations that operate by SOPs; as the Soviets never established nuclear missile bases outside of their country at the time, they assigned the tasks to established departments, which in turn followed their own set procedures. The department’s procedures were designed for Soviet, not Cuban, conditions; hence, mistakes were made that allowed the U.S.A to quite easily learn of the program’s existence. Such mistakes included Soviet troops forgetting to camouflage and even decorating their barracks with Red Army Stars viewable from above.

Model III, the *Governmental Politics Model* is very well suited for answering questions of the form (I) *Why does x have property P, rather than the ideal property P'?* On the whole, being the most fine-grained of the three models, it is probably best suited to serve *therapeutic* or *preventive* I-interest. Thanks to its specificity, Model III allows to describe problems in greater detail and also suggests solutions that, due to their particularity, minimize collateral damage. Additionally, due to the human scale on which it operates, the solutions suggested are easier to implement than in other models; it is easier to fire an incompetent staff member than to change a balance of power. Consider, e.g., the question: *Why did the Soviet Union decide to place offensive missiles in Cuba, rather than not place offensive missiles (and try to improve its bargaining position in another way).* This question emphasizes the actual decision of placing the missiles. From Krushchev’s perspective, closing the missile gap was only one of the options to increase his bargaining position concerning Berlin. It was not the most rational one, because the situation might have led to total annihilation of both sides. To explain this non-ideal action, Model III suggests the path of trying to get a closer understanding of what person Krushchev was and how he looked at the world. Furthermore, Model III emphasizes Krushchev’s personal responsibility and suggests that had someone else been in power, the Cuban Missile Crisis might never have happened.

The crucial point this example from Allison’s classic study illustrates is that we need more than one explanatory model to best answer the different explanation-seeking questions, taking into account the accuracy, adequacy and efficiency of the answers. The three explanatory models provide us with different forms of explanation at different levels and they are indispensable if we want different possible explanation-seeking questions to be answered as good as possible. I do neither claim that one model is always linked to one specific form of questions, nor that

there would never be any competition (or cooperation) among the three models in answering a question of a specific form.<sup>5</sup>

### 9.2.3 *Consequences of the Framework for the Individualism/Holism Debate*

In the remainder of the paper, I elaborate three lessons that this framework for understanding explanatory strategies in social scientific practice can teach us in relation to the individualism/holism debate: (1) we should shift away from debating in terms of one single best form and level of explanation; (2) we should question ontological defences of forms/levels of explanation; and, (3) we should move from a monist mindset to a pluralist mindset advancing explanatory pluralism to go beyond the individualism/holism dichotomy.

Let me start here with lesson (1), shifting the debate away from thinking in terms of one single best form or level of explanation. The plurality of forms of explanation in social scientific practice made social scientists and philosophers discuss about what the best way of explaining a phenomenon would be, often thought of as being *one single* form of explanation (e.g., intentional explanation) or one theoretical perspective (e.g., rational choice theory), and made them advocate that this type or model of explanation had to be implemented in all of the social sciences. I have called this way of discussing the best form of explanation the *winner-takes-all*-approach.

According to this approach, first, it seems that the successfulness of a form of explanation in one particular field, or in relation to one particular question, seems sufficient for many people to claim that it should be used in all possible fields in the social sciences. Second, in the same spirit, it seems that giving one counterexample to a form of explanation that has been favoured as the best one, is enough to discard the form of explanation. Against the winner-takes-all-approach, I defend – in line with the framework presented – that, first, there are *no general preference rules*, i.e. do not expect all answers coming from one and the same account, the same form/level of explanation, and, that, second, there are *no general exclusion rules*, i.e. there are different interests to be addressed, so a form of explanation that fails to answer some explanation-seeking questions convincingly, might perform well on other (current or future) explanation-seeking questions. Thus, we need a rich toolbox with different forms and levels of explanation. Using the framework, one can articulate the strengths and weaknesses of different forms and levels of explanation with respect to different explanatory questions and leave the winner-takes-all-approach behind.

To give you an example of such a winner-takes-all-approach, consider the *fine-grain preference qua explanations* (cf. Elster 1983; Taylor 1988). Two fine-grain

---

<sup>5</sup> I refer the interested reader to our 2007 paper for further details.

preferences can be distinguished: *the small-grain preference* and *the close-grain preference*. In the social sciences, the *small-grain preference* advises to look for detailed individualistic micro-accounts that replace holistic macro-level accounts, like functional and structural explanations. The *close-grain preference* is a matter of favouring explanations that provide the detailed mediating mechanisms in causal chains across time; explanations satisfying this preference will not leave any causal gaps in the temporal chain of events. So any explanatory factor that is at a temporal remove from the fact explained should be replaced by a factor closer, more proximate to the fact, leaving no substantial temporal gaps in the causal chain leading to the event or fact that is explained.

Using the framework of Sect. 9.2.1 in relation to the close grain preference, it can be shown that both remote and proximate causes can be (un)interesting or at least not as interesting as the causal information provided by the other kind of cause (viz. proximate or remote). Thus ignoring remote causes means ignoring possibly important causal information.<sup>6</sup> The small-grain preference deals with the levels of explanation (and is clearly linked to the individualism/holism debate). The discussion turns around whether the best explanations should be found on the individual (lower) level or on the social (higher) level. I have showed how sometimes the lower-level and sometimes the higher-level explanation is the better one.<sup>7</sup> Most defenders of the fine-grain preference neglect the differences in explanatory information between the social higher- and individualistic lower-level explanations, and seem to motivate their fine-grain preference mainly by ontological arguments. While this example of the fine-grain preference focuses on arguments made by methodological individualists, methodological holists also seem to be tempted at times by the idea of there being one single best form of explanation (as will be shown in the example of Lloyd in Sect. 9.3.1), rather than acknowledging the strengths and weaknesses of different forms and levels of explanations – be they individualist or holist.

### 9.3 Emergence, Reduction and Microfoundations in Light of Explanatory Strategies

Equipped with the framework for understanding explanatory pluralism and its consequences, let us now revisit some of the central topics in the individualism/holism debate, i.e. emergence, reduction and microfoundations (their importance in the debate is discussed by Zahle 2006). I evaluate the role emergence plays in contemporary versions of methodological holism in Sect. 9.3.1. In Sect. 9.3.2, I discuss reduction, pondering whether methodological holists have been investing too much in arguing *against* methodological individualism at the cost of losing

---

<sup>6</sup> Cf. Van Bouwel and Weber (2002a).

<sup>7</sup> Cf. Weber and Van Bouwel (2002).

positive aspects of the reductionist drive out of sight. Sometimes, in relation to some explanation-seeking questions, it is very useful to make abstraction of, or distort, the higher-level and apply an individualist explanatory strategy. The approach I want to promote in this paper focuses on the strengths and limitations of the respective explanatory strategies of holists and individualists, which differs from the ways of reasoning that have dominated the individualism/holism debate as identified in this paper – that is also why the title of this paper talks about going *beyond* the current debate. In Sect. 9.3.3, I discuss one form of rapprochement between individualists and holists, i.e. the microfoundations requirement, but a couple of questions remain – questions that will be dealt with in Sect. 9.4.

### 9.3.1 Emergence

Keith Sawyer is one of the researchers that put *emergence* and the use of other concepts of philosophy of mind high on the agenda in the philosophy of social science. Emergence would help to conceptualize the relation between individual and society and to defend higher-level causation and explanation (cf., Sawyer 2001, 2002, 2003). Even though Sawyer acknowledges the “error of making ontological arguments in support of methodological claims” (Sawyer 2002: 538), most of his attention goes to ontological questions when discussing explanation in social science.<sup>8</sup>

Sawyer’s main concern is to develop a full-blown metaphysical picture in order to legitimize social explanations (while sometimes mixing up ontology and methodology). It is important here to explicate the difference in approach between on the one hand Sawyer on emergence (and many other philosophers in the debate) and on the other hand the approach I advocate. As done in Sect. 9.1, it should be highlighted that the individualism/holism debate involves an ontological part and a methodological part (as well as a semantic, legal, ethical, political, etc. part which we will not discuss here). The framework for understanding different explanatory strategies I presented above focuses and intervenes on the methodological part of the debate. A *methodological* approach should be differentiated from an *ontological* approach.

The ontological approach in this debate on explanations is a way of reasoning that starts from arguments about ontological composition to draw methodological conclusions about the best form of explanation. It is present in the fine-grain preference, discussed in Sect. 9.2, and in other defences of a single best form of explanation; the disputed *winner-takes-all* approach is also fed by this ontological approach. A recent example of this approach is given by Pierre Demeulenaere in his introduction to *Analytical Sociology and Social Mechanisms* when he writes that methodological individualism, “can be expressed very simply: Social life exists

---

<sup>8</sup> See Van Bouwel (2004a, 2010) for details.

only by virtue of actors who live it; Consequently a social fact of any kind must be explained by direct reference to the actions of its constituents.” (2011: 4) Just like many philosophers of social science have done before him, Demeulenaere makes an inference from composition to explanation; one starts with certain *a priori* or necessary truths concerning social ontology, the nature of social reality, thus deciding on the locus of causation, justified by ‘metaphysical commonplace’, political convictions, doubtful transcendental arguments, or even without further argument . . . and, then, the methodological consequences, e.g. the best level of explanation, seem to follow ‘automatically’ from the ontological stance.<sup>9</sup>

A similar way of reasoning seems to be present in the debate on emergence. The argument from emergence can be used against individualists to prove that there are important irreducible aspects to be found on the higher, supra-individual level. While my framework could clarify a form of *epistemological emergence* also articulating the indispensability of higher-level explanations, the ontological approach wants to “prove” that in an ontological way, i.e. by claiming *ontological emergence*.<sup>10</sup> The underlying assumption of the ontological approach is that the explanatory should be closely tied to the ontological level: Where individualists go from ontological composition to explanation, emergentists seem to go the other way round, from explanation to composition – the indispensability of higher-level explanations to be anchored in ontological emergence.

These ontological arguments concerning emergence are very prominent in contemporary defences of methodological holism. They seem to be mainly made to prove individualists wrong, but little is said about what it actually implies for explanations (besides the point that there cannot be only individualistic explanations); what are the explanatory restrictions on the basis of ontological emergence (and when does a social constellation lead to emergent properties and when does it not, and how to find out)? What does a satisfactory explanation look like according to defenders of ontological emergence? Does every explanation need some form of *macro-covering* or *macro-roof* (analogous to microfoundations, cf. Sect. 9.3.3. below) according to this contemporary version of methodological holism?

---

<sup>9</sup> I consider drawing explanatory consequences from ontological arguments to be problematic and impoverishing, cf. Van Bouwel and Weber (2002b, 2008b) and Van Bouwel (2004a, c). Notwithstanding the critical questions I raise concerning the ontological approach, let it be clear that ontological debates could play a legitimate role in considering methodological possibilities of a particular theory, model or approach. The ontological moves I criticize are different in that they are often made *a priori* (not on the basis of a thorough study of social scientific practice) and that the results of the ontological statements are to be generalized across the social sciences (not limiting them to the particular theory, model or approach).

<sup>10</sup> *Ontological emergence*, on the one hand, claims that novel, real and irreducible properties do exist (or come into existence) on the higher level. These emergent properties are just as real as physical properties. Following *epistemological emergence*, on the other hand, the concept of emergence is characterized in terms of possibilities of and limits on human knowledge of complex systems: it deals with the (in)adequacy of reducing theories and is based on the fact that it sometimes appears to be impossible to understand the *global* behavior of a complex system by analyzing the *local* behavior of the individual parts. (cf. Van Bouwel 2010)

Looking for answers to these questions, we, first, notice that methodological holists discussing emergence often pay very little attention to the impact on explanation (except for the conclusion that individualist explanations cannot always do the job). Second, for the ones that do the methodological, explanatory consequences often seem to follow ‘automatically’ from the ontological position. (In that sense, the ways of reasoning of some methodological holists and methodological individualists seem to have a lot in common.) One example can be found in Christopher Lloyd’s *ontological and methodological structurism*. Lloyd (1993) identifies a group of social scientists, which can be labelled relationists, that emphasize the linkages between agents and structure and often invoke emergence. Their ontological point of view is labelled the *structurist ontology* by Lloyd (1993: 42–43).<sup>11</sup> But what does this ontology actually imply for our understanding of what a good explanation should look like? Lloyd sketches this in his definition of methodological structurism: “Methodological structurism approaches explanation by developing concepts of the separate real existence yet mutual interdependence of individuals and institutional structures (. . .). Thus methodological structurism is explicitly based on an ontology of the social that recognizes two nodes of causal power.” (Lloyd 1993: 46) The exact implications for what a satisfactory explanation should look like according to structurists are not spelled out, but the thinking in terms of a tight link between ontology and methodology and the lack of considerations about explanatory pluralism and the different explanatory interests resulting in different explanation-seeking questions, is at least obvious.<sup>12</sup> In Sect. 9.4, I return to the issue of relating levels and satisfactory explanations.

Summarizing, much of the discussion concerning emergence in relation to the individualism/holism debate is characterized by the ontological approach (just as it was the case with defences of methodological individualism). This approach tends to unnecessarily restrict (or neglect) the explanatory options available and this leads to a suboptimal situation as concerns the amount of explanation-seeking questions that can be dealt with in the best possible way. That is the second lesson to be drawn from the framework introduced above. Moreover, some of the claims defended by the ontological approach to emergence, could be defended more convincingly by the framework.

### 9.3.2 Reduction

While one can agree with cases of epistemological emergence in answering some explanation-seeking questions about social phenomena, different questions about the same phenomena might be answered best by reductive explanations. While

---

<sup>11</sup> Others have labeled this kind of ontology the *Transformational Model of Social Activity* (TMSA).

<sup>12</sup> For an analysis of structurism and TMSA, see Van Bouwel (2004b, c).

indispensability arguments are normally used to defend higher-level explanations (e.g., Jackson and Pettit 1992), one could also use them to defend lower-level, reductive explanations (see Van Bouwel et al. 2011).

Let it be clear that my defence of reductive explanations does not imply that I support intertheoretic reduction as imperative and a general strategy that always leads to the best and most reliable explanations. Rather, I acknowledge that sometimes explanatory interests are best served accurately, adequately and efficiently by decomposition, by reduction as explanatory strategy. These interests might be more theoretical, e.g., to increase understanding, or more practical, e.g., to find the right level at which to manipulate, change or correct. The focus on one kind of factor can be very productive as scientific practice shows; some causal aspects of a phenomenon might be emphasised, while other aspects might be obscured or perhaps even distorted – pointing once again at the strengths, the partiality and the weaknesses of different explanatory strategies.<sup>13</sup>

When discussing reduction, we should also clarify what level to reduce to. Traditionally in the individualism/holism debate it was presupposed that there be some comprehensive, unique, and privileged individual level (cf. Ylikoski 2012: 26); an individual micro-level which would always be the same level and was contrasted with a macro-level, the social level. However, “more realistic is the understanding that there are social compounds at a range of levels of organization, with different scope and reach” (Little 2012: 138). I do agree with Little and Ylikoski in there being multiple levels of social explanation. Furthermore, the amount and specification of levels is perspectival, depending on the phenomenon at hand. Consider, for instance, the example on the Cuban Missile Crisis in Sect. 9.2.2. and the example on criminal behaviour in a footnote in this section; both use (different) multiple levels in their analysis. These refinements of the traditional dichotomous way of thinking about levels in the individualism/holism debate do not imply that we should stop thinking in terms of higher- and lower-levels, or micro-macro, only that levels are perspectival rather than absolute and unique.

### 9.3.3 *Microfoundations*

A third central discussion I want to revisit briefly, is the one on the microfoundations requirement. This requirement stipulates “that all social facts, social structures, and social causal properties depend ultimately on facts about

---

<sup>13</sup> For those interested, in Van Bouwel et al. (2011), I showed the indispensability of reductive explanations using an example taken from social scientific practice, i.e. comparing the best answers to the following explanation-seeking questions (I) *Why do we have high crime rates in American society?* (II) *Why does criminal or deviant behaviour manifests itself in American-born students A, B, C, but not in foreign-born X, Y, Z?* (III) *Why does person A manifest criminal behaviour, while B does not (even though A and B share the same social environment)?*



individuals within socially defined circumstances. Social ascriptions require microfoundations at the level of individuals in concrete social relationships.” (Little 2012: 138) To discuss this issue, one often uses the *Coleman boat* representing the microfoundations of a macro-level fact (cf. Coleman 1990: 8). The boat visualizes a relationship between macro-factors (e.g., Protestant religious doctrine and capitalism) as well as the micro-factors that underlie this relationship (e.g., values and economic behaviour). It also includes macro-to-micro and micro-to-macro connections.

What does the microfoundations requirement imply for an explanation to be satisfactory? Advocates of the microfoundations approach have been formulating different answers to that question. For some, a macro-explanation (i.e. the upper part of the Coleman boat relating macro-factors like the Protestant religious doctrine and capitalism) will never be satisfactory. For instance, Hedström and Swedberg state in their presentation of the social mechanisms approach: “In the social sciences, however, the elementary “causal agents” are always individual actors, and intelligible social science explanations should always include explicit references to the causes and consequences of their actions.” (Hedström and Swedberg 1998: 11–12) For them, this does not only imply that we need a micro-level part in every explanation, but also: “that there exist no macro-level mechanisms; macro-level entities or events are linked to one another via combinations of situational, individual action, and transformation mechanisms, i.e., all macro-level change should be conceptualized in terms of three separate transitions (macro-micro, micro-micro, and micro-macro).” (Hedström and Swedberg 1996: 299)

As becomes clear after reading this quote, excluding (macro-to-macro) mechanisms on a macro-level does not mean that the defenders of social mechanisms want to exclude *all* references to entities on the macro-level from social explanations. They just consider a reference to (individual actions on) the individual, micro-level as a condition *sine qua non* of a satisfactory explanation. Underlying this claim seems to be an ontological conviction, a conviction concerning causation, namely that causal agents are always individual actors.

Where some see only the lower-part of the Coleman boat as a satisfactory explanation (macro-to-micro, micro-to-micro, micro-to-macro), others see the satisfactory explanation as integrating both macro-level and the micro-level in the explanation. A third option is to consider the macro-level explanation as satisfactory with the condition that an account of the lower part of the boat can be provided (without the latter having to be part of the explanation); microfoundations would play a justificatory rather than an explanatory role. This option is defended by, e.g., Ylikoski (2012) and Little (2012: 143): “The requirement of microfoundations is not a requirement on explanation; it does not require that our explanations proceed through the microfoundational level. Rather, it is a condition that must be satisfied on *prima facie* grounds, prior to offering the explanation.(. . .) In short, we are not obliged to trace out the struts of Coleman’s boat in order to provide a satisfactory macro- or meso-level explanation or mechanism.” Little adds that one argument for his claim is scientific practice itself, “the fact that good sociologists do in fact make credible use of such claims.” (Little 2012: 145) These recent papers of Little and

Ylikoski – defending a microfoundations approach that is different from the earlier ideas of, e.g., Hedström and Swedberg (1996) – bring the idea of microfoundations closer to the spirit of my framework, both by seriously taking into account the actual explanatory practice of social scientists and by avoiding ontological fallacies. A couple of questions remain though.

One might ask, for instance, whether we are not providing all of the time all kinds of explanations and causal claims, without knowing the underlying mechanisms or foundations (cf. Kincaid 1997)? Why would we have this requirement specifically for *social* explanations? What does this requirement mean in practice; when are microfoundations satisfactorily stipulated in order for a macro-explanation to be satisfactory? These are questions that remained to be answered by defenders of the microfoundations requirement. Little links his own account of explanation to pluralism, “this implies the legitimacy of a fairly broad conception of methodological pluralism in the social sciences, constrained always by the requirement of microfoundations.” (Little 2012: 146) It is to be seen to what extent the microfoundations requirement leads to pluralism, or to what kind of pluralism, and is it any different from the pluralism defended by me in Sect. 9.2?

## 9.4 Understandings of Explanatory Pluralism Beyond the Individualism/Holism Dichotomy

Bringing the elements of Sect. 9.3 together, we notice that rather than dichotomous individualism/holism thinking, the focus of recent contributions has been shifting towards interaction, relationism and pluralism, away from the winner-takes-all approach. The question then becomes how to combine individualism and holism, how can they co-exist, interact, be integrated or develop some division of labour? Thus, the debate shifts to how pluralism should be understood.

### 9.4.1 *Different Understandings of Explanatory Pluralism*

The visualization of the Coleman boat and structurism (or the Transformational Model of Social Action), integrating both levels, seem to share the intuitions of Sandra Mitchell’s *integrative pluralism*. Integrative pluralism takes into account both today’s highly specialized (sub)disciplinary research and the need of integrating the respective findings concerning a phenomenon. “Developing models of single causal components, such as the effects of genetic variation, or of single-level interactions, such as the operation of selection on individuals (. . .) need to be integrated in order to understand what historical, proximal, and interactive processes generate the array of biological phenomena we observe. Both the ontology and the representation of complex systems recommend adopting a stance of

integrative pluralism, not only in biology, but in general.” (Mitchell 2004: 81). However complex, and however many contributing causes participated, there is only one causal history that, in fact, has generated a phenomenon to be explained. Thus, according to Mitchell’s integrative pluralism, “it is only by integration of the multiple levels and multiple causes (...) that satisfactory explanations can be generated.” (Mitchell and Dietrich 2006: S78)

Mitchell opposes her integrative pluralism to *isolationist pluralism* or “*levels of analysis*” pluralism. According to that understanding of explanatory pluralism different questions invoke different explanatory schemata, and there is no need to consider explanations developed at levels other than their own or for intertheory relations among the levels. This limits the interaction between various theories offering explanations in a given domain and leads to isolation, according to Mitchell. “If there is no competition between levels, there need be no interaction among scientists working at different levels either.” (Mitchell 2004: 85)<sup>14</sup>

There is (at least) one possible understanding of pluralism which Mitchell does not discuss. Let us label it *interactive pluralism*. It is situated in between integrative and isolationist pluralism, as: (a) on the one hand, it claims that satisfactory explanations can also be obtained without having done the integration of multiple levels, so there is no integration imperative, and, (b) on the other hand, it does not discourage interaction as, in some instances, interaction and integration do lead to better explanations.<sup>15</sup>

#### 9.4.2 *Evaluating Different Understandings of Explanatory Pluralism*

Spelling out different possible understandings of explanatory pluralism beyond the individualism/holism dichotomy, raises the question of which understanding of pluralism is the more convincing one, if any? Below, I briefly raise some challenges concerning integrative and isolationist pluralism, and emphasize the benefits of interactive pluralism.<sup>16</sup>

---

<sup>14</sup> Further, Mitchell (2009) also distinguishes *Anything Goes pluralism* and *moderate pluralism*. The former speaks for itself, the latter is an understanding of pluralism that promotes a temporary plurality of competing theories as a means toward achieving a unified theory in the long run. I will not discuss these forms of pluralism here.

<sup>15</sup> For more taxonomies of pluralism, see, e.g., Kellert et al. (2006) and Van Bouwel (2009).

<sup>16</sup> For a more extensive discussion and evaluation of different understandings of pluralism, also see Van Bouwel (2009).

### 9.4.2.1 Integrative Pluralism

A first challenge concerns whether integration is always necessary to obtain a ‘satisfactory explanation’? Straightforward reduction might sometimes lead to very satisfactory explanations efficiently serving our explanatory interest (cf. Sect. 9.3.2). Integration might very well be a good heuristic advice or play a justificatory role, but why should it be a criterion for a satisfactory explanation?

Second, won’t integrated explanations often provide us with too much information and therefore be less efficient in providing the answers we are looking for. In his book *The rise and fall of the biopsychosocial model*, Nassir Ghaemi (2010), discusses how this model (for psychiatry) included the idea that adding and integrating “more perspectives is always better”. Eventually the approach was made unfeasible in practice by being too general and too vague. Integrative pluralism insufficiently acknowledges that explanations are always a trade-off between generality and preciseness, simplicity and realism, accuracy and adequacy, etc., depending on one’s explanatory interests (cf. Sect. 9.2.1). Integrative explanations might be sometimes far too cumbersome, less efficient, and less adequate than possible alternative explanations. Returning to the example in Sect. 9.2.2, adding up or integrating Allison’s three models (if at all possible) would not lead to better explanations, because it would mean using models (being part of the integration) to satisfy explanatory interests they are not suited for.

Third, integrated explanations might lead to losing *idioms/adequacy* in light of our explanatory interests, thus losing the capacity of answering some explanation-seeking questions in the most adequate way (*i.a.* strengthening hermeneutical injustice).

Fourth, in relation to the third worry, what would the integration imperative imply for *heterodox*, non-mainstream theories? What is the impact on the dynamics between research approaches? Think in particular about situations in which there is epistemic inequality, in which one research program at one level is a lot bigger and more elaborated than another one at another level and where integration risks minimizing dissent, overlooking diversity, eliminating differences and/or a homogenisation in terms of the bigger one. Would the integration imperative then not boil down to adjusting to the mainstream?

### 9.4.2.2 Isolationist Pluralism

A first question: Does isolation always lead to better explanations? And, second, how to know given the lack of competition between explanations within this understanding of pluralism? According to Mitchell’s characterisation of this position, the idea that some questions are better answered on one level and others on another leads to an isolationist stance with respect to the separate questions. Now, if there is no interaction or no intention of competition between levels, then there need be no interaction among scientists working at different levels either. Thus, this form

of pluralism does not do much more than recognizing plurality; it does not suggest any way of making the plurality epistemically as productive as possible. Revisiting the example in Sect. 9.2.2, Allison himself provides us with three different models but without any instruction of how the three models relate to each other or should be used. Without further instruction or framework they just each in isolation give an explanation of the Cuban Missile Crisis at a different level.

Third, why do isolationist pluralists presuppose interaction cannot be productive while fruitful interactions between (sub)disciplines have characterized much of the history of science?

Fourth, as concerns the dynamics between research approaches, isolation, lack of engagement between the orthodoxy and heterodoxy, e.g. in economics, seems to create a very static, non-productive situation in which, on the one hand, the traditional heterodoxy is aiming to become the new monism, the new mainstream substituting the current orthodox one, while, on the other hand, the current orthodoxy or mainstream sees the traditional heterodoxy serving as a constitutive outsider for its own scientificness (cf. Van Bouwel 2009).

### 9.4.2.3 Interactive Pluralism

Interactive pluralism, the possibility not discussed by Mitchell, might be a third option that avoids some of the worries about integrative and isolationist pluralism. Why?

First, where there is a presumption of reconcilability in integrative pluralism, and irreconcilability in isolationist pluralism, interactive pluralism considers the ir-/reconcilability to be an open question, up for interaction.

Second, interactive pluralism questions whether integration would always lead to a better explanation as well as whether integration is necessary to obtain a 'satisfactory explanation'. As concerns the former, integrative explanations might sometimes be too general, vague and cumbersome, i.e. not always the most efficient. Mitchell does not take into account the adequacy and efficiency criteria in stipulating what is the most satisfactory explanation. As concerns the latter claim that integration would be necessary to obtain a satisfactory explanation, I defended above that we should rather consider the trade-off between accuracy, adequacy and efficiency of explanations in labelling what is 'satisfactory'. Always focusing on integration, irrespective of one's precise explanatory aims and needs in a given context, would – if even possible – unnecessarily complicate matters and even paralyze research and decision-making.

Third, even though integration is not imperative, interactive pluralism rejects isolation and endorses interaction and engagement, be it without the presumption of always reaching a consensus or an integration. The respective explanation-seeking questions can be channels of interaction between competing research programs. Going back to Allison in Sect. 9.2.2, I demonstrated how we choose the model that best serves the epistemic interests as made explicit in the explanation-seeking question. Some (but definitely not all!) explanation-seeking questions might require

a combination, integration or cooperation of models, e.g. Model II and III, in order to address our explanatory interests as good as possible.

Fourth, contrary to integrative pluralism, the mainstream and non-mainstream approaches start on equal footing. But even for heterodox approaches that cannot be easily integrated, the interaction with orthodox or other heterodox approaches is endorsed, because approaches are sharpened as a response to challenge and criticism, methodologies refined, concepts clarified, etc. Moreover, the interaction between explanatory approaches might also make the limitations of each approach evident by the articulation of questions that they are not designed to answer.

### 9.4.3 *What's Next?*

The discussion of different understandings of explanatory pluralism above points at a direction in which the individualism/holism debate might move, namely away from a winner-takes-all debate to a debate about how the different approaches should be combined, related or interact. Integrative pluralism might be in line with the ideas of many of the emergentists and microfoundations advocates, while the interactive pluralism fits well with the framework I have presented in Sect. 9.2. In scientific practice, many scientists might call themselves pluralists and rather think in terms of isolationist pluralists or as monists tolerating, but not engaging with, competing approaches. There might very well be plurality in the social sciences, while most social scientists still have a monist mindset.

If you take (one of) the pleas for pluralism of the more active kind (i.e. integrative and/or interactive pluralism) as convincing, then, first, we should continue discussing which understanding of pluralism is the most productive in social science. Second, we should reflect on how that understanding of pluralism can be operationalised and promoted. How can we shape a pluralist mindset among social scientists? How to structure the interaction among competing research approaches in practice? The framework for explanatory pluralism I presented is a possible tool to pay more attention to plurality, a tool by which strengths and weaknesses can be articulated and the winner-takes-all monist mindset can be left behind. Moving beyond the monistic mindset was the third lesson I wanted to draw from the framework for understanding explanatory strategies in social scientific research. It also seems to imply to move beyond the dichotomous individualism/holism debate to a debate about pluralism.

## 9.5 Conclusion: What to Debate About?

Starting from the plurality of explanatory strategies in scientific practice, I presented a framework for explanatory pluralism – as a normative endorsement of plurality – to go beyond the individualism/holism debate on explanation. Rather

than aiming for (or imposing) a full-blown metaphysical picture, I seek to optimize practice; evaluating research approaches against one another, not against a monistic ideal of a single complete and comprehensive account.

Having developed a framework that can be used as a tool for articulating the strengths and weaknesses of different forms and levels of explanation with respect to different explanatory questions, we can leave the winner-takes-all-approach behind, question the ontological defences of the best form/level of explanations, and start moving from a monist mindset to a pluralist mindset.

The next step in the debate is not about developing the ultimate individualistic approach or defending the holist approach, but rather about understanding how different approaches can interact, co-exist, integrate, develop a division of labour, . . . – the pluralism question. That is an actual problem that many social scientists face in their practice and try to find a solution for, for instance, in international political economy (Phillips and Weaver 2011), international relations (Sil and Katzenstein 2010) and in heterodox economics (see Van Bouwel 2004c). I hope my clarifications of the different explanatory strategies beyond the individualism/holism debate might be useful for those social scientists.

**Acknowledgments** The author would like to thank Julie Zahle, Finn Collin and Linnéa Arvidsson for their very helpful comments on earlier versions of this paper.

## References

- Allison, G. (1971). *Essence of decision: Explaining the Cuban missile crisis*. Boston: Little, Brown and Company.
- Allison, G., & Zelikow, P. (1999). *Essence of decision: Explaining the Cuban missile crisis*. New York: Longman.
- Bhargava, R. (1992). *Individualism in social science: Forms and limits of a methodology*. Oxford: Clarendon.
- Bunge, M. (2000). Ten modes of individualism – none of which works – and their alternatives. *Philosophy of the Social Sciences*, 30(3), 384–406.
- Coleman, J. (1990). *Foundations of social theory*. Cambridge, MA: Belknap.
- De Langhe, R., Weber E., Van Bouwel J. (2007). A pragmatist approach to the plurality of explanations in International Relations Theory. In *Conference paper archive of SGIR 6th Pan-European conference on international relations*. <http://www.sgir.eu/conference-paper-archive/>
- Demeulenaere, P. (2011). Introduction. In P. Demeulenaere (Ed.), *Analytical sociology and social mechanisms* (pp. 1–30). Cambridge: Cambridge University Press.
- Elster, J. (1983). *Explaining technical change: A case study in the philosophy of science*. Cambridge: Cambridge University Press.
- Garfinkel, A. (1981). *Forms of explanation: Rethinking the questions in social theory*. New Haven: Yale University Press.
- Ghaemi, N. (2010). *The rise and fall of the biopsychosocial model*. Baltimore: The Johns Hopkins University Press.
- Hedström, P., & Swedberg, R. (1996). Social mechanisms. *Acta Sociologica*, 39, 281–308.
- Hedström, P., & Swedberg, R. (Eds.). (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge: Cambridge University Press.

- Jackson, F., & Pettit, P. (1992). In defense of explanatory ecumenism. *Economics and Philosophy*, 8, 1–21.
- Kellert, S., Longino, H., & Waters, K. (Eds.). (2006). *Scientific pluralism: Minnesota studies in the philosophy of science XIX*. Minneapolis: University of Minnesota Press.
- Kincaid, H. (1997). *Individualism and the unity of science*. Lanham: Rowman & Littlefield.
- Kincaid, H. (2012). How should philosophy of social science proceed? *Metascience*, 21, 391–394.
- Little, D. (2012). Explanatory autonomy and Coleman's boat. *Theoria*, 74, 137–151.
- Lloyd, C. (1993). *The structures of history*. Oxford: Blackwell Publishers.
- Mitchell, S. (2004). Why integrative pluralism? *Emergence Complexity and Organization*, 6(1–2), 81–91.
- Mitchell, S. (2009). *Unsimple truths: Science, complexity, and policy*. Chicago: University of Chicago Press.
- Mitchell, S., & Dietrich, M. (2006). Integration without unification: An argument for pluralism in the biological sciences. *The American Naturalist*, 168, S73–S79.
- Pettit, P. (1993). *The common mind: An essay on psychology, society and politics*. New York: Oxford University Press.
- Phillips, N., & Weaver, C. (Eds.). (2011). *International political economy: Debating the past, present and future*. New York: Routledge.
- Risjord, M. (2000). *Woodcutters and witchcraft: Rationality and interpretive change in the social sciences*. Albany: State University of New York Press.
- Sawyer, R. K. (2001). Emergence in sociology: Contemporary philosophy of mind and some implications for sociological theory. *American Journal of Sociology*, 107(3), 551–585.
- Sawyer, R. K. (2002). Nonreductive individualism. Part I – Supervenience and wild disjunction. *Philosophy of the Social Sciences*, 32, 537–559.
- Sawyer, R. K. (2003). Nonreductive individualism. Part II – Social causation. *Philosophy of the Social Sciences*, 33, 203–224.
- Searle, J. (2009). Language and social ontology. In C. Mantzavinos (Ed.), *Philosophy of the social sciences: Philosophical theory and scientific practice* (pp. 9–27). Cambridge: Cambridge University Press.
- Sil, R., & Katzenstein, P. (2010). *Beyond paradigms: Analytic eclecticism in the study of world politics*. Basingstoke: Palgrave MacMillan.
- Taylor, M. (1988). Rationality and revolutionary collective action. In M. Taylor (Ed.), *Rationality and revolution* (pp. 63–97). Cambridge: Cambridge University Press.
- Udehn, L. (2001). *Methodological individualism: Background, history and meaning*. London: Routledge.
- Van Bouwel, J. (2003). *Verklaringspluralisme in de sociale wetenschappen*. Ph.D. thesis, Ghent University.
- Van Bouwel, J. (2004a). Individualism and holism, reduction and pluralism. *Philosophy of the Social Sciences*, 34, 527–535.
- Van Bouwel, J. (2004b). Questioning structurism as a new standard for social scientific explanations. *Graduate Journal of Social Science*, 1, 204–226.
- Van Bouwel, J. (2004c). Explanatory pluralism in economics: Against the mainstream? *Philosophical Explorations*, 7, 299–315.
- Van Bouwel, J. (2009). The problem with(out) consensus: The scientific consensus, deliberative democracy and agonistic pluralism. In J. Van Bouwel (Ed.), *The social sciences and democracy* (pp. 121–142). Basingstoke: Palgrave Macmillan.
- Van Bouwel, J. (2010). Why social emergence? Discussing the use of analytical metaphysics in social theory. In R. Vanderbeeken & B. D'Hooghe (Eds.), *Worldviews, science and us: Studies of analytical metaphysics* (pp. 153–167). Singapore: World Scientific.
- Van Bouwel, J., & Weber, E. (2002a). Remote causes, bad explanations? *Journal for the Theory of Social Behaviour*, 32, 437–449.
- Van Bouwel, J., & Weber, E. (2002b). The living apart together relationship of causation and explanation. *Philosophy of the Social Sciences*, 32(4), 560–569.



- Van Bouwel, J., & Weber, E. (2008a). A pragmatic defense of non-relativistic explanatory pluralism in history and social science. *History and Theory*, 47, 168–182.
- Van Bouwel, J., & Weber, E. (2008b). De-ontologizing the debate on social explanations: A pragmatic approach based on epistemic interests. *Human Studies*, 31(4), 423–444.
- Van Bouwel, J., Weber, E., & De Vreese, L. (2011). Indispensability arguments in favour of reductive explanations. *Journal for General Philosophy of Science*, 42(1), 33–46.
- Van Fraassen, B. C. (1980). *The scientific image*. Oxford: Clarendon.
- Weber, E., & Van Bouwel, J. (2002). Can we dispense with the structural explanation of social facts? *Economics and Philosophy*, 18, 259–275.
- Ylikoski, P. (2012). Micro, macro, and mechanism. In H. Kincaid (Ed.), *The Oxford handbook of philosophy of social science* (pp. 21–45). Oxford: Oxford University Press.
- Zahle, J. 2006. Holism and Supervenience. In S. Turner & M. Risjord (Eds.), *Handbook of Philosophy of Anthropology and Sociology* (pp. 311–341). Amsterdam: Elsevier.

# Chapter 10

## Holism, Emergence, and the Crucial Distinction

Julie Zahle

**Abstract** One issue of dispute between methodological individualists and methodological holists is whether holist explanations are dispensable in the sense that individualist explanations are able to do their explanatory job. Methodological individualists say they are, whereas methodological holists deny this. In the first part of the paper, I discuss Elder-Vass' version of an influential argument in support of methodological holism, the argument from emergence. I argue that methodological individualists should reject it: The argument relies on a distinction between individualist and holist explanations that they find unacceptable and Elder-Vass' reasons in support of his way of drawing this distinction are not good ones. In the second part, I examine what, if anything, would be good reasons in support of a particular way of differentiating between individualist and holist explanations. I propose that a good reason is one which shows, in an acceptable manner, that the distinction, drawn in the same way in all contexts, is useful from the perspective of offering explanations of the social world. I show that if this criterion is adopted, it will result in a fruitful reorientation of the debate between methodological individualists and methodological holists.

### 10.1 Introduction

The methodological individualism-holism debate covers several issues. One important issue is whether holist explanations are dispensable in the sense that individualist explanations are able to do their explanatory job. Roughly speaking, methodological individualists hold that individualist explanations, that is,

---

J. Zahle (✉)

Department of Media, Cognition and Communication, University of Copenhagen,  
Copenhagen, Denmark  
e-mail: [jzahle@hum.ku.dk](mailto:jzahle@hum.ku.dk)

explanations in terms of only individuals, their actions, and the like, can do all the explanatory work needed. By contrast, methodological holists insist that holist explanations, that is, explanations in terms of social wholes, their actions, and the like, must *also* be invoked in order not to leave important events unexplained.

Since the late 1970s, a number of arguments have been put forward in support of methodological holism. A central argument is the argument from emergence: By appeal to the notion of emergence, it is claimed that unless holist explanations are sometimes offered, it is impossible to account for significant events in the social world. The argument has been presented in different versions by social scientists and philosophers alike. Within the social sciences, its advancement is often associated with theorists working within the highly influential school of Critical Realism. Leading critical realists such as Margaret Archer, Roy Bhaskar, Andrew Collier, Dave Elder-Vass, Tony Lawson, and Andrew Sayer have all offered different versions of this argument.<sup>1</sup> The present paper is centered round an examination of Elder-Vass' recent advance of the argument in "The Causal Power of Social Structures" (Elder-Vass 2010). The book is among the best in the critical realist tradition and it contains a particularly strong version of the argument from emergence.

The discussion falls in two parts. In the first and main part of the paper, I critically discuss whether Elder-Vass' argument from emergence should make methodological individualists give up their claim that holist explanations are dispensable. I begin by outlining Elder-Vass' argument. Next, I show that current methodological individualists tend to draw the line between individualist and holist explanations differently than Elder-Vass. Accordingly, they may dismiss his argument on the ground that it relies on an unacceptably narrow conception of individualist explanations. In response, I suggest, Elder-Vass might try to argue that current methodological individualists *should* adopt his distinction between individualist and holist explanations. Thus, I proceed to examine Elder-Vass' reasons in support of his distinction. I show that these reasons are not good ones. On this basis, I conclude that, as it stands, Elder-Vass' argument from emergence should not make current methodological individualists abandon their position.

In the second part of the paper, I discuss the more general question of whether it is even possible to defend a particular way of drawing the crucial distinction between individualist and holist explanations. I propose that a good reason in support of a particular distinction is one which shows, in an acceptable manner, that the distinction, drawn in the same way in all contexts, is useful from the perspective of explaining the social world. We should start examining – and evaluating – particular distinctions as to whether they are supported by good

---

<sup>1</sup> See, e.g., Archer (1995, 2000), Bhaskar (1982, 2000[1979]), Collier (1989, 1994), Elder-Vass (2007, 2010, 2012), Lawson (1997), and Sayer (2010) for critical realist discussions of the argument from emergence. It should be stressed that, very often, critical realists do not themselves present their arguments of emergence as arguments in favor of methodological holism. This is because they use the term in a different sense than I do. In view of the definition stated above, their arguments are straightforward arguments in support of methodological holism.

reasons in this sense. This is rarely, if ever, done. Yet, I show, it holds the key to a fruitful reorientation of the dispute between methodological individualists and holists.

## 10.2 Elder-Vass' Argument from Emergence

Elder-Vass' argument from emergence draws heavily on ontological considerations. Accordingly, it is helpful first to provide a rather detailed presentation of his account of the structure of reality. On that basis, I outline his notion of explanation and his distinction between individualist and holist explanations. Finally, I state his version of the argument from emergence.

Elder-Vass' account of the structure of reality takes as its starting point the notion of an entity. As Elder-Vass defines it, an entity is a relatively enduring whole composed of parts standing in certain relations to each other (Elder-Vass 2010: 17). Of course, these parts may be entities too and hence they may themselves be persistent wholes made up of parts standing in certain relations to each other, and so on. As an illustration of this idea consider biological organisms. These are wholes composed of parts in the form of cells or molecules related in certain ways and, in turn, these parts are made up of parts in the form of atoms standing in certain relations to each other (ibid.: 66). Perhaps in the process of unraveling reality in this manner, parts may be found that are not themselves composed of yet further parts. Disregarding this limiting case, however, reality may be seen as containing various entities both combining and dissolving into yet other entities.

Elder-Vass elaborates on this ontological account by pointing out that an entity may have various properties. Consider an entity which forms part of a whole. From the perspective of its functioning as a part, the entity has various non-emergent properties, as I shall call them. These are the properties of a part which it also has in isolation or as an element in an unstructured collection of parts.<sup>2</sup> The part would also possess these properties even if it were not, at that very moment, part of a given whole made up of structured parts. The property of mass possessed by oxygen and hydrogen atoms as combined to form water exemplifies a non-emergent property in this sense: The mass possessed by each of the atoms is a non-emergent property since each atom would also have a certain mass without being, at that moment in time, a part of a structured whole in the form of water.

Next, turn to an entity from the perspective of its being a whole. Elder-Vass discusses various kinds of properties which a whole may have. To begin with, the resultant or aggregative properties of a whole are those properties of a whole that its parts also have in isolation or as elements in an unstructured collection of parts

---

<sup>2</sup>For the sake of simplicity, I shall often talk about those properties which parts may have in isolation. Thus, I shall leave out the second part "or as an element in an unstructured collection of parts."

(*ibid.*). So for example, the property of having a mass as possessed by water is a resultant property since this property is also possessed by oxygen and hydrogen atoms in isolation. The resultant properties of wholes differ markedly from their emergent ones. The latter are the properties of a whole that its parts do *not* have in isolation and “that would not be possessed by the full set of parts in the absence of a structuring set of relations between them” (*ibid.*). The emergent properties of a whole are ones it has in virtue of, or because of, its parts standing, *at that moment*, in certain relations to each other. Here, “at that moment” is meant to signal that the relationship between the emergent properties of a whole and its parts is synchronic and non-causal rather than diachronic and causal (*ibid.*: 23).<sup>3</sup>

Elder-Vass differentiates between two kinds of emergent properties of wholes. One I shall refer to as the emergent properties of wholes proper. These are exemplified by the property of water to be a liquid at certain temperatures: This property is not possessed by its parts, *viz.* hydrogen and oxygen atoms, in isolation (Elder-Vass 2007: 29). The other type of emergent properties of wholes requires a little more explanation: The parts of a whole may have properties that they would not possess in isolation, that is, if they were not part of a structured set of components. For instance, the parts of a star in the form of particles have the property of emitting light and the particles would not have this property unless they were, at that moment, interrelated in a certain way (Elder-Vass 2010: 59). One might be tempted to regard these properties as emergent properties of the parts and I shall sometimes refer to them as such. However, Elder-Vass maintains that a property of this type “is really a property of the whole that happens to be localized in some respect within the part” (*ibid.*: 27). This view may be expressed by saying that these properties are really emergent properties of wholes that have been delegated to their parts (Elder-Vass, personal communication).

In addition to this classification into different kinds of properties, Elder-Vass points out that a property of any type is “some intrinsic aspect of an entity that can have a causal impact on the world” (Elder-Vass 2010: 17). The fact that the aspect must be intrinsic means that an entity being, say, larger than another entity should not be regarded as a property since this relationship between the two entities is purely formal. Because being a property goes together with having the capacity causally to influence the world, Elder-Vass talks interchangeably about properties and (causal) powers.

These are the main elements in Elder-Vass’ account of the general structure of reality. He applies these considerations to a special kind of entities, *viz.* *social* entities. These entities may be characterized as enduring wholes composed of parts

---

<sup>3</sup> This understanding of emergent properties is one among many. For different understandings of the notion of emergence, see, e.g. Beckermann et al. (1992), Bedau and Humphreys (2008), Bunge (2003), Corradini and O’Connor (2010), Kim (1999), and O’Connor and Wong (2012). For an overview of the appeal to emergence within sociology, see Sawyer (2005).

in the form of individuals standing in certain relations to each other.<sup>4</sup> There are different kinds of social entities. For instance, one notable kind is organizations such as firms, schools, universities, religious organizations, households, and the like (see *ibid.*: 144).<sup>5</sup>

Like other entities, social wholes and their parts, viz. individuals, have various properties. Individuals, as parts, have non-emergent properties, that is, properties that they would also possess in isolation. These are exemplified by the power to talk, to eat, to overthrow a vase, and to read.<sup>6</sup> Social entities, as wholes, also have the three types of properties discussed above. They have resultant properties. The power to talk, to make a lot of noise, and to march as ascribed to social wholes, such as a kindergarten or an army, are instances of resultant properties of social entities since their parts, the individuals, would also possess these properties in isolation or as elements in an unstructured collection of parts. Further, social entities have emergent properties proper. For example, a social whole in the form of a pin factory may possess the emergent power to produce more. This qualifies as an emergent property when the factory has this property as a result of the factory workers standing in certain relations to each other: Between them, they have divided up the task of producing pins (*ibid.*: 154). Other examples of these emergent properties might be a country's power to initiate a war, a government's power to introduce a new tax or a barbershop quartet's "ability to produce [. . . a] harmonized performance" (*ibid.*). Finally, social wholes have emergent properties delegated to their parts in the form of individuals. As an illustration of this point, consider an individual's power to fire, to hire, to vote, and to grade. These properties are emergent ones because they exist in virtue of individuals being interrelated such as to form a social whole. They are powers ascribed to individuals but because these powers have been delegated to them by social wholes, they are really properties of social wholes. Henceforth, I shall also refer to these properties as individuals' *part-emergent* or *p-emergent* properties to distinguish them from those properties of individuals which qualify as emergent insofar as individuals are regarded as wholes rather than as parts.

Turning now to Elder-Vass' notion of explanation, his view is that an explanation states how an entity, in virtue of its properties/causal powers, partially brought

---

<sup>4</sup> At one point, Elder-Vass notes that social entities may also have material things as their parts (*ibid.*: 157). Still, he does not make much of this insight and hence I shall also ignore this possible way of spelling out parts of social entities.

<sup>5</sup> In his (2010), Elder-Vass provides a nice and detailed analysis of organizations as well as norm circles, which are another kind of social entities. He extends this analysis to further kinds of social entities in his (2012).

<sup>6</sup> Note that as soon as individuals are regarded not as parts but as wholes, their non-emergent properties may instead qualify as emergent ones, that is, properties which are the result, synchronically speaking, of individuals being themselves composed of parts standing in certain relations to each other. Unless otherwise noted, I have in mind individuals' non-emergent properties *as parts* when talking about these properties.

about an aspect of a specific event.<sup>7</sup> As it stands, this may seem like an account of an incomplete explanation: All it conveys is how *one* aspect of an event was *partially* brought about. However, as Elder-Vass explicates, an explanation can only focus on one out of many aspects of a particular event just as it can only mention some – the most significant – of the multiple causal powers that resulted in the occurrence of a given aspect of an event (ibid.: 178).

Elder-Vass does not spend much effort on spelling out the difference between individualist and holist explanations. Nor does he provide a lot of examples of the two sorts of explanation. Still, in light of his various comments, it is reasonable to hold that he takes an individualist explanation to specify how an individual or individuals, in virtue of her/their properties, partially brought about an aspect of a particular event. Thus, some clear instances of individualist explanations might be that the vase fell down because Amy knocked it over; that there was no more food in the camp because the girls had eaten the rest of it; and that Simon was quickly done because he was a fast reader. These explanations count as individualist because it is individuals who bring about various aspects of events in virtue of their non-emergent powers exemplified by the power to knock something over, to eat something, and to read fast.

By contrast, Elder-Vass maintains that a holist explanation states how a social entity or entities, in virtue of its/their properties, partially brought about an aspect of a particular event. Examples of these explanations are that Ben does not have a job anymore because the manager of the firm fired him (see ibid.: 74); that the owner of the pin factory earned more money than the year before because the factory produced more pins; and that Simon couldn't help her because the battalion kept on marching. These explanations are holist because it is social entities which bring about various aspects of events in virtue of their emergent and resultant properties, namely the power to fire, to produce more pins, and to march. It is true that the account of why Ben doesn't have a job anymore states that it is an individual, viz. the manager, who fires him. Still, the explanation should be regarded as a disguised or masked holist explanation since it is *really* a social entity, viz. a firm, that has the emergent power to fire Ben and that exercises this power through the manager.

In light of all these clarifications, Elder-Vass' version of the argument from emergence may now be presented. Here is how it may be reconstructed: In order for holist explanations to be dispensable, it must be possible to substitute them with individualist explanations. More specifically, it must be the case that whenever an aspect of a particular event is said to have been partially brought about by social wholes, the same aspect of that event may equally well be said to have been partially brought about by individuals.

---

<sup>7</sup> Elder-Vass works with two different notions of explanation: Explanations in the above causal sense and what may be called reductive explanations. I shall not be concerned with reductive explanations and hence I use "explanation" to refer to his causal notion of explanation only. On reductive explanations, see, e.g., Elder-Vass (2010: 24).

Holist explanations that appeal to the resultant properties of social wholes are dispensable in this sense. The resultant property of a social whole is also possessed by its parts, viz. individuals, in isolation: The property may be ascribed to individuals as a non-emergent property. An effect of a resultant property possessed by a social whole is nothing over and above the effect of the summation of the corresponding non-emergent property possessed by the individuals who compose the whole. Therefore, this effect may equally well be said to have been partially brought about by this set of individuals in virtue of their non-emergent properties. In this fashion, individualist explanations may do the explanatory job of these holist explanations.

Holist explanations that refer to the emergent properties of social wholes is a different matter. These properties are not possessed by their parts, viz. individuals, in isolation. Instead, they are properties that social wholes have in virtue of their parts, individuals, standing in certain relation to each other. This means that an effect of an emergent property possessed by a social whole “is something more than the effects that would be produced by the entity’s parts [viz. the individuals] if they were not organised into this sort of whole” (ibid.: 66). An effect of an emergent property possessed by a social whole *is* something over and above the effect of individuals in virtue of their non-emergent properties. As a result, an effect produced by a social whole in virtue of its emergent property may not equally well be said to have been partially brought about by its parts, viz. individuals, in virtue of their non-emergent properties. Since individualist explanations may only state how individuals, in virtue of their properties, partially brought about an aspect of a specific event, they cannot perform the explanatory job carried out by holist explanations. Add to these considerations that holist explanations are indeed advanced within the social sciences, and it follows that their use is indispensable. This amounts to a vindication of the thesis of methodological holism.

### **10.3 Elder-Vass’ Distinction Between Individualist and Holist Explanations**

The argument from emergence purports to show that holist explanations are indispensable. The question is, however, whether the argument should make methodological individualists give up their position. I examine this issue by focusing on whether methodological individualists are likely to accept an important assumption made by Elder-Vass’ argument.

To set off the discussion, note that methodological individualism comes in various versions and that one main way in which to differentiate between these versions is in terms of how the line between individualist and holist explanations is drawn. Historically speaking, there has been a move among methodological individualists towards still broader conceptions of individualist explanations (Udehn 2002: 498). It is by appeal to increasingly broader specifications of an individualist explanation that methodological individualists have defended the claim that individualist explanations can do all the explanatory work of holist explanations. As



part of this development, current methodological individualism maintains that individualist explanations are allowed to refer to certain relations between individuals<sup>8</sup>:

Individualist explanations involve “statements about the dispositions, beliefs, resources, and *interrelations* of individuals” (Watkins 1973b[1957]: 168 – my italics).

King’s (1999) “position is straightforward methodological individualism: ‘[...] social life [...] can be adequately accounted for by reference only to individuals, their practices and *relations*’ (p. 278)” (Sawyer 2005: 89 – my italics).<sup>9</sup>

To get anywhere, we always have to assume *relations* between individuals, as well as individuals themselves. The works of methodological individualists themselves underline this point (Hodgson 2007: 218 – my italics).

From these quotes, it is not clear exactly what sort of relations the authors have in mind and, not surprisingly, there are diverse ways in which to spell out the notion of a relation. For the present purposes it is enough to register that current methodological individualists typically take “relations between individuals” to refer to many kinds of relations including the sort of relations which go together with, or make it possible, for individuals to perform actions like firing, hiring, voting, and grading. As a result of the latter point, current methodological individualists maintain that explanations which describe individuals as firing, hiring, voting, grading, and the like, should be regarded as individualist explanations.<sup>10</sup> This assertion conflicts with Elder-Vass’ position. From his perspective, individuals exercise their p-emergent powers when they fire, hire, and so on. So, he would rephrase current methodological individualists’ view as the contention that an explanation which states how individuals, in virtue of their p-emergent properties, partially brought about an aspect of an event, should be regarded as an individualist explanation. According to Elder-Vass, this is wrong. These explanations fall within the category of holist explanations and the reason is, remember, that individuals’ p-emergent powers are *really* properties of social wholes delegated to individuals. Thus, explanations of this form are disguised holist explanations: Unmasked, they state

---

<sup>8</sup> Perhaps there are a few exceptions this trend. However, for ease of exposition and simplicity, I shall ignore these exceptions in the following. I take current methodological individualism to cover defenses of the position going back to the 1950s at least.

<sup>9</sup> For the record, King does not regard his own position as methodological individualist (see his (2007)). Still, this is irrelevant in the present context: What matters is that Sawyer takes the position to be a *clear cut* example of methodological individualism.

<sup>10</sup> This is for instance made clear by Steven Lukes in his classic paper “Methodological Individualism Reconsidered” (Lukes 1995[1968]). Here, Lukes refers to explanations containing descriptions of individuals as voting, hiring, etc. as explanations which contain type (iv) predicates while noticing that the use of these explanations is “extremely widespread” among methodological individualists (ibid.: 455). In the paper, Lukes opposes the use of these predicates by methodological individualists. In another famous paper, Mandelbaum did the same (1973[1955]). Particularly, the latter paper sparked a lot of debate on this issue. See, e.g., Bhargava (1992), Danto (1973[1962]), Gellner (1968), Goldstein (1973[1958], 1959), James (2009[1984]), Lesnoff (1974), Little (1991), Kincaid (1994, 1997), Martin (1972), Quinton (1975), and Zahle (2003, 2007).

how social wholes partially brought about an aspect of an event.<sup>11</sup> In this fashion, the disagreement between current methodological individualists and Elder-Vass concerns the manner in which to partition the territory of explanations, so to speak. Current methodological individualists take more explanations to count as individualist and hence they leave fewer to qualify as holist than Elder-Vass does. Current methodological individualists advocate a broad conception of individualist explanations compared to Elder-Vass' narrow one.

The difference of opinion has consequences when it comes to current methodological individualists' assessment of Elder-Vass' argument from emergence: They may simply point out that the argument relies on an assumption which they deny, viz. that individualist explanations should be narrowly circumscribed. In fact, they may stress, the argument importantly trades on this presupposition since, roughly speaking, the narrower the conception of individualist explanations, the easier it is to show that these cannot do all the explanatory work of holist explanations. Since they do not subscribe to a narrow conception of individualist explanations, methodological individualists may continue, they reject the argument.<sup>12</sup>

Of course, there are various ways in which Elder-Vass might react to this response. One option would be to make an attempt to defend the argument by arguing that current methodological individualists *ought* to adopt his distinction between individualist and holist explanations. Evidently, if Elder-Vass managed to show this, current methodological individualists could no longer reject his argument on the ground that they disagree with the manner in which he draws this distinction. In what follows, I explore this option. I concentrate on Elder-Vass' defense of the claim that explanations should be categorized as holist rather than individualist when they state how individuals, in virtue of their p-emergent properties/powers, partially brought about an aspect of a particular event. Methodological individualists may well want to dispute Elder-Vass' way of phrasing these explanations. When this matters for the argumentation, I shall refer to them as explanations that refer to individuals as firing, hiring, voting, etc. Otherwise, I shall simply stick to Elder-Vass' manner of formulating the explanations.

---

<sup>11</sup> Notice that it really matters to current methodological individualists whether these explanations are classified as individualist ones: It is only if they are regarded as individualist explanations that current methodological individualists maintain that it is possible to defend their position that *all* holist explanations are dispensable.

<sup>12</sup> This sort of response, i.e. the dismissal of arguments on the grounds that the specification of individualist and holist explanations is unacceptable, is not uncommon in the debate, see, e.g., Mäki (2002) and Watkins (1973c[1955]).

## 10.4 Elder-Vass' Defense of His Distinction

In support of his preferred way of drawing the distinction between individualist and holist explanations, Elder-Vass appeals to ontological considerations. Here is how his reasoning may be summarized: The distinction between individualist and holist explanations should be drawn such that it *maps onto* a more basic ontological distinction between social entities, as wholes, and individuals, as parts. Following this ontological distinction, individuals' p-emergent properties are *really* properties of social entities delegated to individuals. By implication, explanations which refer to these properties should be regarded as disguised holist explanations: Unmasked, they state how social wholes, in virtue of their properties, partially brought about an aspect of a specific event.

At first sight, the basic idea informing this claim may seem puzzling: How is this suggestion of mapping to be understood? I begin by explicating this idea by introducing the terminology of levels. Once formulated in terms of levels, Elder-Vass' claim turns out to exemplify a common way of thinking.

Both within the philosophical and scientific literature, it is quite common to talk about levels.<sup>13</sup> Various kinds of things are sorted into levels. These include explanations, theories, sciences, entities, properties, and vocabularies. The relationship between things at different levels is hierarchical (Kim 2002: 4). Accordingly, if explanations, say, are sorted into levels, then some explanations should be seen as being at a higher – or lower – level than others.

Further, levels in different senses, that is, levels applied to different kinds of things, are often taken to go together or map onto each other (Craver 2007: 172ff). This way of thinking is nicely illustrated by Paul Oppenheim and Hilary Putnam in their classic paper “Unity of Science as a Working Hypothesis” (Oppenheim and Putnam 1958). Here, they make it clear that objects, vocabularies, theories, and sciences may be sorted into levels such that the objects at each level line up with a distinct science with its distinct level of vocabulary and theories (see Craver 2007: 172ff). Or, to mention a more recent example, Bermudéz describes the idea that sciences at different levels offer explanations at different levels: “it is often believed that the natural sciences slot together to give a unified, multi-level explanatory picture of the physical world. Many philosophers, psychologists and cognitive scientists think that we need to see the different disciplines as operating at different levels of the hierarchy, offering explanations that complement rather than compete with each other” (Bermudez 2005: 28).

Frequently, mapping claims are taken one step further. It is made clear that a level in one sense is regarded as prior, or more basic, than levels in other senses. Craver may be interpreted as giving expression to this view. He proposes “that we start by thinking of levels as primarily features of the world” while continuing in the

---

<sup>13</sup> Very informative discussions of levels talk may be found in, e.g., Craver (2007), Kim (2002) and Wimsatt (1994).

footnote that “[l]evels of sciences and theories could then be seen as derivative upon, and at best approximations of, these ontic structures” (Craver 2007: 177).

Against this background, Elder-Vass’ mapping claim may be formulated in terms of levels. As explained earlier, Elder-Vass stresses that social entities are wholes which have individuals as their parts. This being the case, he submits, social entities should be regarded as higher-level entities and individuals as lower-level entities (see Elder-Vass 2010: 19). Likewise, the emergent and resultant properties of social entities should be seen as higher-level properties, whereas the non-emergent properties of individuals should be classified as lower-level properties. Also, in line with the tradition for thinking about levels, Elder-Vass takes this ontological distinction to line up with, or go together with another distinction, namely a distinction between explanations: Holist explanations – or higher-level explanations – state how social entities (=higher-level entities), in virtue of their resultant and emergent properties (=higher-level properties) partially brought about aspects of particular events. And individualist explanations – lower-level explanations – state how individuals (=lower-level entities) in virtue of their non-emergent properties (=lower-level properties) partially brought about aspects of particular events.

Finally, Elder-Vass thinks that his ontological distinction is basic while considering his distinction between individualist/lower-level and holist/higher-level explanations as derivative upon it. This is for instance seen by the fact that Elder-Vass’ (2010) – the main focus of this paper – is centered round his ontological account which involves his discussion of social entities, as wholes and higher-level entities, and of individuals as parts and lower-level entities. These reflections are then used as basis for his sketchier and scattered remarks about issues of explanation. Also, it is made clear by his comment that his methodological recommendations, including his view on how social scientists should approach the task of explaining the social world, should be regarded as “a means of putting into practice” his account of the general structure of reality (ibid.: 65). By presenting Elder-Vass’ mapping claim in this fashion, I think, it appears that it is quite common. In light of the tradition of levels talk, there is nothing unusual about his idea that the distinction between holist/higher-level and individualist/lower-level explanations should be drawn such that it *maps onto* the ontological distinction between social/higher-level entities and individuals/lower-level entities.

Be that as it may, Elder-Vass maintains that his ontological part/whole distinction involves a classification of individuals’ p-emergent properties as really properties of social entities as wholes. Call this Elder-Vass’ preferred version of the ontological part/whole distinction. There is an obvious alternative to it. The alternative view has it that individuals’ p-emergent properties are really their own properties. Elder-Vass’ preferred ontological distinction forms the basis for his claim that explanations which appeal to these properties should be regarded as disguised holist explanations. Conversely, if the alternative version is adopted and individuals’ p-emergent properties are regarded as really properties of individuals, then explanations that refer to these properties meet the definition of individualist explanations. These state, remember, how individuals, in virtue of their properties,

partially brought about an aspect of an event. In defense of his distinction between individualist and holist explanations, Elder-Vass' efforts are directed at showing that his preferred version of the ontological part/whole distinction should be adopted. More precisely, he offers two considerations in support of his claim that individuals' p-emergent properties are really properties of social entities.

In a number of places, Elder-Vass states that the "entities that are H's [a whole's] parts would not have this causal power [p-emergent property] if they were not organised into an H, hence it is a causal power of H and not of the parts" (ibid.: 24). In relation to social entities, this translates into the claim that since individuals would not have their p-emergent powers if they were not parts of social entities, these powers are really the powers of the social entities and not of the individuals. As Elder-Vass explains it, a "manager could not dismiss an employee unless both were parts of an organization of a certain kind, thus the causal power is a power of the organization, exercised on its behalf by the manager, and not a power of the manager as an individual" (ibid.: 74).

When reasoning in this manner, Elder-Vass presupposes the acceptance of a specific criterion of property ascription. He assumes that when deciding how to classify the emergent properties of parts, the decisive consideration is whether the part would also have the property in isolation: If it wouldn't, then the power is really a power of the whole. But obviously it is equally possible to hold that the decisive consideration is, say, whether a part is actually ascribed the emergent power and that, if it is, then this power should be regarded as the part's alone. Or, it may be maintained that the crucial consideration is whether the part may implement or exercise the emergent power, and that if so, then the part alone should be said to possess this power. Insofar as these alternative criteria are used, the p-emergent properties of individuals qualify as being really their properties and not really the properties of social entities. Since there are various possible criteria that one might adopt, it is reasonable to demand that Elder-Vass should say something more as to why his criterion should be espoused. This is particularly so since its acceptance would mean a change in our current practices: As also recognized by Elder-Vass, we tend to oppose the idea that individuals' p-emergent powers are really the properties of social entities (see ibid.: 198–199).

Elder-Vass' second consideration takes this last point into account. Consider these statements:

When I mark an exam paper, for example, and assign a mark, I do so as an arm of my university, not as an individual [...] it is the university that has the power to assign a mark and I do so on its behalf. Just as when I use my own arm to lift a bag it is me (as well as my arm) that is doing the lifting, so in the marking case it is the university (as well as me) that is doing the marking (ibid.: 199).

[I]t is the organisation that sells the television, though it does so through the sales assistant who is one of its parts [...] Just as we accept that human beings are causally responsible for the behaviour of their parts when it is directed by their decisions, so we must accept that organizations are causally responsible for the behaviour of their members or employees when that behaviour is motivated by organisational policy and roles (ibid.: 173).

As a first shot, the gist of these and similar comments may be summarized as follows: When a body part of an individual does something, in virtue of an emergent power that it has as a part, *and* when that body part is being directed by the individual's decision, then we regard the emergent power as being really a property, not of the body part, but of the individual, as a whole. To be more consistent in our practices, we should adopt the same approach to individuals' p-emergent properties. Thus, when an individual, as part of a social entity, does something, in virtue of an emergent power that it has as a part, and when the individual is, at the same time, directed by the social entity's organizational policy and roles, then we should also regard the p-emergent power as being really a property, not of the individual, but of the social entity as a whole.

As it stands, this argument will not do. It states that when individuals act in virtue of p-emergent powers *while being also* directed by the social entity's organizational policy and roles, then these powers should really be ascribed to the social entity. However, there are clearly situations in which individuals act in virtue of their p-emergent powers *without* being directed by the social entity's organizational policy and roles. For instance, assume that the male boss in a firm coerces a female employee to have sex with him and that his power to do so is an emergent power: He would not have this power without being the boss of the firm. Also, make the reasonable assumption that when the boss coerces the woman to have sex with him, he is not motivated by the firm's policies or roles. In relation to cases like these, Elder-Vass' reflections do not establish that individuals' p-emergent powers to act should really be ascribed to social entities. In other words, the argument has too limited a scope: It only shows that *some* of the p-emergent properties of individuals should be regarded as being really properties of social entities.

One way to get around this objection is to modify the argument such that it leaves out any reference to body parts and individuals as being directed by decisions and organizational policies respectively.<sup>14</sup> Yet, I am not sure this helps much. Assume that Elder-Vass is right that we regard the emergent properties of body parts as being really the properties of individuals as wholes. Compare this to some other examples of how we classify emergent properties of parts. Think of windscreen wipers as parts of a car. Their power to wipe the windows is an emergent one: They would not have this power unless they were properly wired up to other parts of the car. Here, I think, we tend to say that the wipers, and not the car, have the causal power to wipe the windows. Or, contemplate a spark plug as part of an engine. Its power to emit a spark is also an emergent one: Without being properly wired up to other parts of the engine, it would not have this power. Again, I think, we tend to regard the power to emit a spark as being really a power of the spark plug and not of the engine. If this is right, then there are various other contexts in which we don't classify the emergent properties of parts as being really the properties of their wholes. It is not only when it comes to the p-emergent properties of individuals that we are inclined to oppose their categorization as really properties

---

<sup>14</sup> Thanks to Elder-Vass (personal communication) for pointing me to this reading of his view.

of social entities as wholes. But then why insist that we should change our classificatory practice when it comes to individuals' p-emergent properties in order to bring it in line with how we classify the emergent properties of body parts? In the absence of an answer to this question, Elder-Vass' second consideration is not really persuasive. On a more general note, therefore, he does not succeed in defending his preferred version of the ontological part/whole distinction against the alternative version. He has not persuasively demonstrated that individuals' p-emergent properties should be regarded as being really properties of social entities and not of individuals. Accordingly, he has not provided good reasons for holding that explanations which refer to individuals' p-emergent properties should be regarded as disguised holist explanations.

There is also another problem with Elder-Vass' line of arguing. Above I spent some time spelling out Elder-Vass' idea that the distinction between individualist and holist explanations *should* be drawn such that it maps onto the ontological part-whole distinction. But *why should* this be so? For instance, inspired by John W. N. Watkins' work, it might instead be submitted that the distinction between explanations should map onto an ontological division between individuals as so called moving forces in history versus social entities as lacking this force (see Watkins 1973b[1957]: 168). Or, it might be held that the distinction between individualist and holist explanations should line up with the distinction between explanations based on more accessible evidence versus explanations based on less accessible evidence (see Watkins 1973a[1952]: 150, 165). In relation to both these propositions, it might be argued, explanations which contain descriptions of individuals as firing, hiring, voting, and so on, will qualify as individualist. For the present purposes, the merit of these suggestions is irrelevant. The important point to note is that Elder-Vass does not defend his position against views which, on the one hand, take explanations that refer to individuals' p-emergent properties to count as individualist while, on the other hand, also denying that the distinction between individualist and holist explanations should map onto an ontological part-whole distinction. In effect, Elder-Vass provides no reasons for holding that there *should* be such a mapping relation. In this respect, too, Elder-Vass' defense of his distinction between individualist and holist explanations must be found wanting.

The upshot of all these considerations is that the reasons Elder-Vass offers in support of his favorite distinction between individualist and holist explanations are not good ones. Hence, it is hard to see why current methodological individualists should adopt his distinction. This being the case, current methodological individualists should hold on to their claim that the argument from emergence, or at least Elder-Vass' version of it, may be rejected on the ground that it significantly trades on an assumption that they find unacceptable: It presupposes that individualist explanations are very narrowly circumscribed. By implication, the argument should not make current methodological individualists give up their thesis that individualist explanations may do all the explanatory work of holist explanations.

## 10.5 What Would Be Good Reasons in Support of a Distinction?

In the preceding section, I have discussed Elder-Vass' defense of his preferred distinction between individualist and holist explanations. Now, I examine the more general question of what, if anything, would be good reasons in support of a particular distinction between individualist and holist explanations.

An examination of this question is a natural corollary to the assessment of Elder-Vass' considerations in support of his distinction: All through the previous discussion one may have wondered whether it is even possible to provide good grounds for drawing the distinction between individualist and holist explanations in a given manner. Also – and importantly – there is another way in which to motivate reflection upon this issue. The disagreement about the manner in which to circumscribe individualist and holist explanations does not only exist between Elder-Vass and current methodological individualists. Within the debate more generally, there are numerous views on how this crucial distinction should be spelled out. Moreover, like Elder-Vass, some participants in the debate are careful to offer reasons in defense of their favorite distinction, just as they present objections to those distinctions which they find unacceptable. Among other things, it has been widely discussed whether, when individuals are described as voting, firing, and hiring, this should count as a holist or as an individualist explanation. Likewise, it has been debated whether individualist explanations should be allowed to describe particular individuals only or whether they should also be permitted to describe individuals as types, and so on. Yet, as Susan James also observes, the problem of the right way to differentiate between individualist and holist explanations “has proved surprisingly resistant to any quick solution” (James 2009[1984]: 38). Despite all the efforts to settle this issue, not much progress has been made. In view of this situation, it is natural to wonder whether it is even possible to offer good reasons in support of a particular distinction between individualist and holist explanations. It is relevant in relation to the individualism-holism debate more generally to get clear on what, if anything, would constitute good reasons in support of a particular distinction.

As a first step in the direction of an answer, note a rather trivial, but often overlooked, point with respect to distinctions: They should serve a useful purpose.<sup>15</sup> Otherwise, why draw them, let alone insist on them? In relation to distinctions between individualist and holist explanations, there are various useful purposes which they may serve. For instance, a particular distinction may be deemed useful from the perspective of generating fun philosophical discussions or because it goes neatly together with various philosophical commitments. In my view, these suggestions are unsatisfactory: They do not sufficiently take into account that the methodological individualism-holism debate concerns explanations as advanced within the social sciences. Accordingly, it should be specified that

---

<sup>15</sup>This point is also made by Kim in relation to distinctions into levels (Kim 2002).



a particular distinction between individualist and holist explanations must be useful from the perspective of offering social scientific explanations. This means that a particular distinction may be said to be useful because it, say, marks a division between two ways in which to arrive at or confirm explanations, or because it marks a difference in the sort of practical or theoretical interest that we take in these explanations. Irrespective of why a particular distinction is claimed to be useful, it is important that such an assertion is made on an acceptable basis. That is, if it maintained that a particular distinction is useful since it, say, marks two different ways in which explanations should be confirmed, then this latter claim must be correct too. I shall put this point by saying that a distinction must be shown useful in an acceptable manner. Finally, note that participants in the methodological individualism/holism debate regard the distinction between individualist and holist explanations as a fixed or context invariant one: They take it that the distinction should be drawn in the same manner in all contexts.<sup>16</sup> This means that it is a particular *fixed* distinction that must, in an acceptable manner, be shown useful from the perspective of facing the task of explaining social reality.

In light of these considerations, a good reason in support of a particular distinction between individualist and holist explanations may be specified as follows: It is a reason which shows, in an acceptable manner, that the distinction, drawn in the same manner in all contexts, is useful from the perspective of explaining the social world. Call this the pragmatic requirement. Insofar as a particular distinction is not supported by reasons that meet this requirement, I submit, it should be abandoned. Conversely, if it is supported by reasons that meet this criterion, it should be regarded as an option. I think we should require that whenever a distinction between individualist and holist explanations is invoked, its proponent must show that there are pragmatic reasons in favor of it, that is, reasons which live up to the pragmatic criterion. We should ask: In what way may this particular context invariant distinction, in an acceptable way, be shown to be useful from an explanatory point of view?

If this proposal is taken to heart by the participants in the methodological individualism-holism debate, it will result in a reorientation of the whole dispute. The nature of the reasons advanced in support of particular distinctions will radically change: Currently these rarely, if ever, include considerations relating to explanatory usefulness. It is instructive briefly to consider what may happen if we start demanding that particular distinctions must be supported by reasons that live up to the pragmatic requirement. It may turn out that there is only a single distinction that is backed up by pragmatic reasons. In this case, the dispute may continue as hitherto with the only important difference that the disagreement about the right way to draw the crucial distinction would have been solved. Another possibility is that several distinctions between individualist and holist explanations

---

<sup>16</sup> Ylikoski, too, notices in relation to individualist explanations that it is assumed that they should be specified in the same manner in all relevant contexts (Ylikoski 2012: 26). He opposes this presupposition from a different perspective than I do below.

appear to be supported by reasons that fulfill the pragmatic requirement. This may ensue since there are various respects in which distinctions may potentially be useful from an explanatory point of view. Hence, a number of particular distinctions may appear to be useful each in their own way. In this case, all these distinctions should be regarded as options. These options should then, as far as possible, be compared in terms of their explanatory usefulness and the most useful of them should be adopted: This distinction will be the one supported by the best reason(s). Finally, it may happen that no distinction may be justified by appeal to pragmatic reasons. Hence, no distinction would qualify as an option and the methodological individualism-holism debate should be abandoned: It is premised on the idea of there being an interesting fixed distinction between individualist and holist explanations.

Here, I shall not engage in speculation on what is most likely to happen. Instead, I want to point to two further ways in which the espousal of the pragmatic requirement, with its emphasis on explanatory usefulness, may potentially change the terms of the debate. According to the criterion, a good reason is one that shows how a distinction, drawn in the same manner in *all* contexts, is useful from an explanatory perspective. But, of course, it is perfectly possible to imagine that it may be shown, in an acceptable manner, that a particular distinction, drawn in the same manner in *certain* contexts, is useful from an explanatory perspective. In these contexts, it might well be argued, the distinction should be regarded as an option. Why not? This point brings out that a central assumption in the methodological individualism-holism debate, viz. the idea of the distinction being context invariant, is in need of justification: Perhaps it is just more useful, from an explanatory perspective, to work with context variant distinctions between individualist and holist explanations. Another feature of the pragmatic criterion is its focus on the explanatory usefulness of distinctions between *individualist* and *holist* explanations. This aspect of the criterion makes it natural to wonder about the following: What if it is shown, in an acceptable manner, that a context invariant (or variant) distinction into three classes rather than the two classes of individualist and holist explanations is explanatorily useful? For instance, it might be suggested to divide explanations into those that refer to individuals with their non-emergent properties, individuals with their p-emergent properties, and social entities. It might reasonably be held that these sorts of alternative classifications should be regarded as options too if they are explanatorily useful. But, of course, this would constitute a radical break with another assumption, underlying the whole debate, to the effect that there is only one distinction of this sort worth drawing between social scientific explanations, viz. that between individualist and holist explanations. This presupposition, too, stands in need of justification. These pressures against the basic assumptions informing the current methodological individualism-holism debate cannot, I think, be ignored. I am not sure how to withstand them but, once more, I want to refrain from making any predictions as to the outcome of addressing these issues. The aim of the present section has been to show that particular distinctions between individualist and holist explanations may indeed be supported by good reasons specified as ones that meet the pragmatic criterion. I have argued that if we begin to

require that particular distinctions must be backed up by pragmatic reasons, it will result in a fruitful reorientation of the whole debate. The exact result of this reorientation remains to be seen.

## 10.6 Conclusion

In the first part of this paper, I discussed Elder-Vass' version of an important argument in support of methodological holism, the argument from emergence. I argued that current methodological individualists should reject it: The argument relies on a distinction between individualist and holist explanations that they find unacceptable and Elder-Vass does not offer good reasons in support of his manner of drawing the distinction. In the second part, I examined whether it is even possible to present good reasons in support of a particular way of differentiating between individualist and holist explanations. I proposed that a good reason in favor of a particular distinction shows, in an acceptable manner, that the distinction, drawn in the same way in all contexts, is useful from the perspective of explaining the social world. Finally I showed how the adoption of this pragmatic criterion will lead to a fruitful reorientation of the whole debate between methodological individualists and holists.<sup>17</sup>

## References

- Archer, M. S. (1995). *Realist social theory: The morphogenetic approach*. Cambridge: Cambridge University Press.
- Archer, M. S. (2000). For structure: Its reality, properties, and powers: A reply to Anthony King. *The Sociological Review*, 48, 464–472.
- Beckermann, A., Flohr, H., & Kim, J. (Eds.). (1992). *Emergence or reduction?* Berlin: Walter de Gruyter.
- Bedau, M. A., & Humphreys, P. (Eds.). (2008). *Emergence. Contemporary readings in philosophy and science*. Cambridge, MA: The MIT Press.
- Bermúdez, J. L. (2005). *Philosophy of psychology*. New York: Routledge.
- Bhargava, R. (1992). *Individualism in social science*. Oxford: Clarendon.
- Bhaskar, R. (1982). Emergence, explanation, and emancipation. In P. F. Secord (Ed.), *Explaining human behavior* (pp. 275–310). Beverly Hills: Sage.

---

<sup>17</sup> I would like to thank Dave Elder-Vass as well as Jeroen van Bouwel, Finn Collin, Tuukka Kaidesoja, Klemens Kappel, Andras Szigeti, and Petri Ylikoski for their very helpful comments. Also, I have given versions of this paper as a talk at the workshop Individualism, Holism, Emergence and Explanations, at University of Copenhagen in 2012, and at the Philosophy of Science (POS) Seminar, Finnish Centre of Excellence in the Philosophy of the Social Sciences, at University of Helsinki in 2013. In both cases, I have benefitted greatly from the questions and remarks made by the audiences. Work on this essay was supported by The Danish Council for Independent Research | Humanities (FKK).

- Bhaskar, R. (2000[1979]). *The possibility of naturalism* (3rd ed.). London: Routledge.
- Bunge, M. (2003). *Emergence and convergence*. Toronto: University of Toronto Press.
- Collier, A. (1989). *Scientific realism and socialist thought*. Hemel Hempstead: Harvester Wheatsheaf.
- Collier, A. (1994). *Critical realism. An introduction to Roy Bhaskar's philosophy*. London: Verso.
- Corradini, A., & O'Connor, T. (Eds.). (2010). *Emergence in science and philosophy*. New York: Routledge.
- Craver, C. F. (2007). *Explaining the brain*. Oxford: Clarendon.
- Danto, A. C. (1973[1962]). Methodological individualism and methodological socialism. In J. O'Neill (Ed.), *Modes of individualism and collectivism* (pp. 312–338). London: Heinemann.
- Elder-Vass, D. (2007). For emergence: Refining Archer's account of social structure. *Journal for the Theory of Social Behaviour*, 37(1), 25–44.
- Elder-Vass, D. (2010). *The causal power of social structures. Emergence, structure and agency*. Cambridge: Cambridge University Press.
- Elder-Vass, D. (2012). *The reality of social construction*. Cambridge: Cambridge University Press.
- Gellner, E. (1968). Holism versus individualism. In M. Brodbeck (Ed.), *Readings in the philosophy of the social sciences* (pp. 254–268). New York: Macmillan.
- Goldstein, L. J. (1959). Mr. Watkins on the two theses. *British Journal for the Philosophy of Science*, 10(39), 240–241.
- Goldstein, L. J. (1973[1958]). Two theses of methodological individualism. In J. O'Neill (Ed.), *Modes of individualism and collectivism* (pp. 277–286). London: Heinemann.
- Hodgson, G. M. (2007). Meanings of methodological individualism. *Journal of Economic Methodology*, 14(2), 211–226.
- James, S. J. (2009[1984]). *The content of social explanation*. Cambridge: Cambridge University Press.
- Kim, J. (1999). Making sense of emergence. *Philosophical Studies*, 95, 3–36.
- Kim, J. (2002). The layered model: Metaphysical considerations. *Philosophical Explorations*, V(1), 2–20.
- Kincaid, H. (1994). Reduction, explanation and individualism. In M. Martin & L. C. McIntyre (Eds.), *Readings in the philosophy of social science* (pp. 497–513). Cambridge: The MIT Press.
- Kincaid, H. (1997). *Individualism and the unity of science*. Lanham: Rowman & Littlefield.
- King, A. (1999). The impossibility of naturalism: The antinomies of Bhaskar's realism. *Journal for the Theory of Social Behaviour*, 29, 267–288.
- King, A. (2007). Why I am not an individualist. *Journal for the Theory of Social Behaviour*, 37(2), 211–219.
- Lawson, T. (1997). *Economics and reality*. London: Routledge.
- Lesnoff, M. H. (1974). *The structure of social science: A philosophical introduction*. London: George Allen & Unwin.
- Little, D. (1991). *Varieties of social explanation*. Boulder: Westview Press.
- Lukes, S. (1995[1968]). Methodological individualism reconsidered. In M. Martin & L. C. McIntyre (Eds.), *Readings in the philosophy of social science*. Cambridge: The MIT Press.
- Mäki, U. (2002). Explanatory ecumenism and economics imperialism. *Economics and Philosophy*, 18, 235–257.
- Mandelbaum, M. (1973[1955]). Societal facts. In J. O'Neill (Ed.), *Modes of individualism and collectivism* (pp. 221–234). London: Heinemann.
- Martin, M. (1972). Explanation in social science: Some recent work. *Philosophy of the Social Sciences*, 2(1), 61–81.
- O'Connor, Y., & Wong, H. Y. (2012). Emergent properties. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/archives/spr2012/entries/properties-emergent/>
- Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Concepts, theories, and the mind-body problem, Minnesota studies in the philosophy of science II* (pp. 3–36). Minneapolis: University of Minnesota Press.

- Quinton, A. (1975). Social objects. *Proceedings of the Aristotelian Society*, 75, 1–27.
- Sawyer, R. K. (2005). *Social emergence. Societies as complex systems*. Cambridge: Cambridge University Press.
- Sayer, A. (2010). *Method in social science. A realist approach*. London: Routledge.
- Udehn, L. (2002). The changing face of methodological individualism. *Annual Review of Sociology*, 28, 479–507.
- Watkins, J. W. N. (1973a[1952]). Ideal types and historical explanation. In J. O'Neill (Ed.), *Modes of individualism and collectivism* (pp. 143–165). London: Heinemann.
- Watkins, J. W. N. (1973b[1957]). Historical explanation in the social sciences. In J. O'Neill (Ed.), *Modes of individualism and collectivism* (pp. 166–178). London: Heinemann.
- Watkins, J. W. N. (1973c[1955]). Methodological individualism: A reply. In J. O'Neill (Ed.), *Modes of individualism and collectivism* (pp. 179–184). London: Heinemann.
- Wimsatt, W. C. (1994). The ontology of complex systems: Levels, perspectives, and causal thickets. *Canadian Journal of Philosophy*, 20(Suppl.), 207–274.
- Ylikoski, P. (2012). Micro, macro, and mechanisms. In H. Kincaid (Ed.), *The Oxford handbook of philosophy of social science*. Oxford: Oxford University Press.
- Zahle, J. (2003). The individualism-holism debate on intertheoretic reduction and the argument from multiple realization. *Philosophy of the Social Sciences*, 33(1), 77–99.
- Zahle, J. (2007). Holism and supervenience. In S. P. Turner & M. Risjord (Eds.), *Handbook of the philosophy of science. Philosophy of anthropology and sociology* (pp. 311–341). Amsterdam: Elsevier.

# Chapter 11

## Who Are the Agents? Actor Network Theory, Methodological Individualism, and Reduction

Finn Collin

**Abstract** This chapter examines one of the most influential developments in current social theory, namely Actor Network Theory (ANT), in the perspective of the individualism-holism issue. ANT moves the actor to the centre of the social stage, yet an actor different from anything that has been encountered before in the literature. The development of the basic ideas of ANT is traced as these ideas gradually evolved in the work of its chief protagonist, Bruno Latour. Latour's thought has a distinctively philosophical flavour, and yet his ideas have been put into practice with ostensible success in a particular area of social research, namely the field of science and technology studies. Latour's work is still in progress, but its current version incorporates many ideas that are also encountered in other recent authors, among them several represented in the current volume. These ideas, however, are pushed beyond their normal bounds, which leads to a dissolution of the classical conception of methodological individualism and results in a radical position that is dubbed "methodological particularism". An examination of this position holds important methodological lessons for social science.

### 11.1 Introduction

Actor Network Theory (ANT) is one of the most influential developments in current social theory. With great fanfare, it moves the actor to the centre of the social stage, yet an actor different from anything we have encountered before. In the same movement, the long fought-over issue of holism vs. individualism in social science is dramatically recast, both from the point of view of ontology and methodology.

---

F. Collin (✉)

Department of Media, Cognition and Communication, University of Copenhagen, Copenhagen, Denmark

e-mail: [collin@hum.ku.dk](mailto:collin@hum.ku.dk)

There are radically new ideas at play here, some of them with a distinctive philosophical tint, and yet they have been put into practice with ostensible success in a particular area of social research: the fast-growing field of science and technology studies. Hence, an examination of ANT could not fail to hold methodological lessons for social science – both positive and negative. In particular, such an examination will be instructive with respect to the problems of methodological individualism vs. holism and of micro-macro reduction.

Although ANT today constitutes a broad academic movement and has received theoretical input from a number of people, one person stands out as the central source of the basic ideas, namely Bruno Latour. The theory bears the mark of his highly inventive mind and his penchant for radical thought, expressed in a correspondingly innovative literary form heavily infused with metaphor. A philosopher as well as a social scientist by training, Latour has authored a string of original works in which substantial sociological insights and methodological reflections are fused in the manner with which we are familiar from the great historical figures of social science. Never since Durkheim and Weber has there been a social scientist so concerned with methodology, and for whom this concern is so intimately interwoven with his scientific endeavours. Hence, Latour's work will be in focus in the following pages.

The many strands of Latour's thought come together in his monograph *Reassembling the Social* (Latour 2005), to which considerable attention will be devoted in this chapter. I shall start my account with Latour's early writings, however, and give a chronological account, necessarily brief and compressed, of the development of his work. For it is in the *dynamics* of Latour's thought that his methodological concerns are primarily manifest.

At the outset, Latour's project was the narrower one of *explaining natural science in social terms*, following in the footsteps of the Strong Programme in the sociology of science. This is *prima facie* one of the most challenging tasks for sociology, attempting to include into the compass of sociology a social institution that, by traditional understanding, is effectively insulated against societal influence.

In his endeavour to socialize science, Latour found himself forced to deal with ever smaller units as the ontological basis of his analysis, thus gradually descending along the micro-macro scale. Latour moved from macro-sociology to micro-sociology and anthropology, only to end, not in methodological individualism, but something even more radical which I shall name *methodological particularism*. In adopting this position, Latour also moves into the borderland between social science and metaphysics.

Thus, Latour's work exhibits, in a highly instructive manner, the problems of individualism vs holism and of reduction in social science, with scientific practice serving as the object of demonstration. Analyzing his work in this perspective will serve to throw light upon such principled issues as explanation and reduction.

## 11.2 ANT and the Individualism-Holism Issue

One of the chief motives behind the push towards individualism in general is what I shall call the *explanation argument*. It has several strands, all of them clearly in evidence in Latour's work.

The social sciences are traditionally regarded as operative on a number of levels. In economics, this stratification is firmly institutionalized in terms of a division into macroeconomics and microeconomics, while sociology sometimes offers a slightly more complex picture with its tri-partition into macro-, meso- and micro-levels. This division represents a methodological embarrassment to the social sciences, since it threatens them with perennial division and fracture. More specifically, it prevents sociology from developing a uniform vocabulary of theoretical concepts applicable across the board, condemning it to balkanization into different specialties, the interconnection of which remains moot.

In the history of sociology, attempted solutions to this problem have moved in different directions. One is the "way upwards", where the micro- and meso-levels are absorbed into the macro-level. An example is Marxist social theory, where the individual is considered to be an abstraction from a social totality, another is Talcott Parsons's "Grand Theory", where the individual agent is absorbed into general societal functions. However, the move downwards to the individual level has been much more popular. This is driven by powerful intuitions of an ontological and explanation-theoretical kind favouring micro-analysis over macro-abstraction. One crucial intuition is that macro-level entities are impotent in an explanatory role, *inter alia* with respect to individual actions. Societal macro-facts are generally thought to be *supervenient* upon, and thus determined by, societal micro-facts. Suppose we explain a particular instance of individual-level social interaction at time  $t$  (call it  $I_t$ ) by reference to the causal influence of a prior macro-societal event (call it  $M_{t-n}$ ). Since  $M_{t-n}$  is by hypothesis supervenient upon some simultaneous individual-level event ( $I_{t-n}$ ), it seems we can eliminate the supervenient event from the equation:  $I_t$  is determined, and hence explained, by  $I_{t-n}$ , and we may disregard  $M_{t-n}$ , which turns out to be a mere epiphenomenon.

Intuitively, the direction of explanation goes the opposite way. Events at the macro-level are explicable in micro-social terms. More precisely, it is a matter of explanation *plus* determination via supervenience between the macro- and micro-levels, such that a complex micro-state causes – and hence explains – a subsequent micro-state, which, by supervenience, fixes a certain macro-state that is thus indirectly explained. An example: A sum of individual employee actions of asking for a pay increase elicits (that is, causes) a sum of actions by employers of granting these demands; more money is subsequently paid out in salaries. Add the background circumstance that the value of production remains constant in the meantime (plus a few other factors), and what you get, by a relation of supervenience, is the macro-level fact of an inflation, but explained in terms of the individual demands.

The downward direction of analysis is further favoured by the perceived need for *mechanisms* making up the links along which the causal forces are channelled.



This requirement is akin to the principle of “No action at a distance”, a bone of contention in seventeenth century physics. Critics objected to the idea, supposedly inherent in Newtonian mechanics, that celestial bodies could affect each other across vast distances of empty space; instead, they insisted that concrete mechanisms of the push-and-pull variety must be involved (a celebrated example is Descartes’s vortex theory). In a similar manner, methodological individualists in social science dismiss talk of individual action being determined by “social forces” reaching across the ontological divide separating the macro- and the micro-levels. This gap can only be traversed by the operation of concrete mechanisms, with the inevitable consequence that the macro-entities involved are broken down into complexes of micro-items. Otherwise, the explanation will fail to link the macro-items referred to in the *explanans* with the micro-items occurring in the *explanandum*. Hence, the demand for mechanisms favours the micro-analytic perspective.

All these motives are clearly in evidence in Latour’s writings, in particular his most explicitly methodological treatise, *Reassembling the Social* (Latour 2005). Now the micro-analytic approach would normally result in *methodological individualism*, according to which the activities of individual human actors serve as the bottom level in terms of which social states and processes are explained. This is often accompanied by an ontological, reductivist claim, to the effect that society is nothing “over and above” the interactions among human individuals. The latter, however, is a position strongly rejected by Latour. He does not consider the individual to be the natural stopping point of micro-analysis in social science, but proceeds to subject even this level to analytic unpacking. As a matter of fact, Latour rejects the very idea of singling out any particular level at all as privileged, from the point of view of social explanation. Latour is an arch-enemy of reductionism, in social science as elsewhere, as signalled by the title of one of his works, which is simply named *Irreductions* (Part 2 of Latour 1988a).

Two brief caveats before we begin our examination of Latour’s development. First, while the micro-macro issue offers a fruitful perspective for an examination of Latour’s work, it does not exhaust it nor lay bare all its sources. Latour is a highly complex thinker, operating in the borderland between social science and philosophy, and there are many further concerns motivating his position.

Secondly, like many other French *maître penseurs*, Latour resents being pigeonholed into categories, in particular such as have become established within Anglo-American academia, seeing his own work as transcending all such traditional categories. Thus, predictably, he dismisses the terms of the traditional micro-macro debate (Latour 2005, p. 169 ff. p. 180). We should note this, but not be deterred by it; I hope to show that analyzing Latour’s work according to classical categories of analytic philosophy of science is actually a worthwhile project that throws light on both sides. (Parts of the following discussion of Latour’s work are based upon my exposition in Collin 2011).

### 11.3 The Point of Departure: *Laboratory Life*

In the work that first catapulted him into international celebrity, *Laboratory Life* (with Steve Woolgar, 1979, 2nd ed. 1986), Latour starts out by paying his dues to the Strong Programme, the pioneering effort within science studies. He presents his own work as a continuation and extension of that programme: “By maintaining the sense in which we use social, we hope to be able to pursue the strong programme at a level apparently beyond traditional sociological grasp” (Latour and Woolgar 1979/1986, p. 152; also pp. 105–107).

At the outset, the approach shared the Strong Programme’s ambition of *explaining* the emergence of scientific theories, and with the same rationale: By explaining the development of science in purely social terms, we can demonstrate that science is not exceptional, but indeed social through and through. Latour and Woolgar argue that the social approach to science can close all gaps in its explanation, in particular such that refer to “intimate thought processes” of a rational nature that supposedly resist social reconstruction (*op. cit.*, p. 168).

The Strong Programme was firmly lodged within a macro-sociological framework, with Marx and Durkheim as its main inspirations (Bloor 1976/1991). From Durkheim, it inherited the idea that the members of society, including scientists, grasp the world in categories that are derived from society and somehow mirror it. We describe nature in ways that project social reality onto it. From Marx, it adopted the idea that the selection among such culturally engendered ideas, of which there will always be a plurality, is dictated by class interests: The dominant ideas will be the ideas of the dominant class.

The overall theoretical framework in *Laboratory Life* derives from economics, in the form of an updated “political economy” of clearly Marxist provenance (*op. cit.*, Chapter 5). The scientific enterprise is viewed as an analogue to, or even an aspect of, the economic cycle of capitalist production. And indeed, scientific activities may bring rich economic rewards to those who excel in them. The basic commodity in science is not money, however, but *credibility*, and all exchanges serve to strengthen it. The model is the Marxist analysis of capitalist economy as governed by the principle of the accumulation of capital: Science, like the material economy, is propelled by the principle of growth. As the authors put it, “Consequently, there is no ultimate goal to scientific investment other than the continual redeployment of accumulated resources. It is in this sense that we liken scientists’ credibility to a cycle of capital investment” (*op. cit.*, p. 198).

This is a macroscopic approach that depicts the familiar characteristics of science as reflecting its institutional structure rather than the individual thought processes of its participants. It is characteristic of what we call science that new and better theories continually replace earlier ones; this is often attributed to a particular critical, rational mindset in its practitioners. But in Latour’s view, we should simply see this as a structural property of the scientific institution, in particular its mechanisms for distributing credibility as the reward for scientific excellence, a prize to be gained in an open, competitive process. Then, we only need to add the almost

truistic premise that human behaviour is guided by the prospect of reward to set the system in motion. Like in the “real” economy, the result is a fierce rivalry among individuals and what Schumpeter called “the creative destruction of capital”. The counterpart of this in science is the way that new theories replace old ones, as a result of the constant competition among scientists for prominence and professional credibility. In both spheres, the secret lies in a macro-structural property – a market, or quasi-market – which makes for a dynamics that we traditionally construe as progress.

Already in this early work, however, Latour and Woolgar took a first step away from Edinburgh orthodoxy and its inherent macro-orientation, descending to the level of detailed societal micro-explanations where they would undertake an “anthropology of science” (*op. cit.*, p. 27 f). Their method would be the classical ethnographic device of participant observation, with Latour entering the Salk laboratories at La Jolla, California, to record what goes on in a biochemistry laboratory.

What motivates this move is a methodological intuition of a micro-analytic kind – more precisely, the desire for *mechanisms*, combined with the urge to demonstrate the “ordinariness” of science. It is not enough to point to certain structural macro-features of science as the source of its apparent rationality. Latour and Woolgar want to show how these features emerge out of the micro-level of human interaction; specifically, that the supposed “rationality” of the individual scientists plays no role in these interactions.

Latour and Woolgar’s adoption of anthropological method does not imply that they will “go native”, however; they will not employ the terms with which scientists themselves describe their own enterprise. The authors adopt an “irreverent” approach to science, refusing to accept its authority in the same way as they would refuse to “bow before the knowledge of a primitive sorcerer” (*op. cit.*, p. 29). They consider scientists’ discourse about science as serving to veil the true character of the scientific process (that is, its status as a human artifact), just as magical practices are designed to conceal that nothing out of the ordinary is really taking place and that only human dexterity is in play. Magical practices are meant to make it appear as if spirits are at work, without human intervention; similarly, science is conducted in a way to make it appear that an independent nature reveals itself directly to us, without intermediaries.

Thus, Latour and Woolgar invite us on an anthropological journey to visit the strange tribe of neuro-endocrinologists, to determine the nature and purpose of their practices. What emerges is the surprising finding that laboratory science is basically about the production and circulation of *texts* with the aid of various “inscription devices”, pieces of apparatus that can transform matter into written documents (*op. cit.*, p. 51). Examples are mass spectrophotometers, radio-immunoassays, and high pressure liquid chromatographs, among others. These are pieces of measuring apparatus the immediate products of which are graphs or numbers; this output is later transformed into discursive texts, which are then published as input to the “credibility cycle”, the operation of which is the very essence of science, as we learned from the initial macro-sociological picture.

Like other representatives of Science and Technology Studies (STS), Latour and Woolgar want to bring out the contingent and a-rational nature of this process (*op. cit.*, p. 152). There is no need to refer to rational thought-processes to explain the outcome of the proceedings, for the very good reason that, strictly speaking, such processes do not exist, but are mere fictions generated by the way scientific research is reported (*op. cit.*, p. 171). This fiction serves obvious strategic ends on the part of the scientific profession, but the anthropologist of science has the power to look “through the veil” and discern the real micro-sociological processes at work.

The “anthropological observer” concludes that laboratory activity, and hence the institution of science to which that activity is central, is the “organisation of persuasion through literary inscription” (*op. cit.*, p. 88). A key characteristic of the scientific process, from an anthropological perspective, is the way that it hides its own nature from those engaged in it, deriving its efficacy precisely from this concealment. For science to work, it must appear as if its products are *not* the result of persuasion, but rather of supra-personal rational principles.

With the adoption of ethnographic methods, Latour and Woolgar take a step away from macro-sociological approaches based upon such traditional categories as “structure”, “class”, “power”, or “interest”. Still, the authors insist that their approach is a social one; they add, however, that they use the term “social” in a somewhat unorthodox sense, referring to what they call the “processual” aspect of science (*op. cit.*, p. 32). In any case, their approach might fairly be called “micro-sociological”, since it deals with the dynamics of small groups and the role that negotiations play therein.

## 11.4 Limitations of the Early Approach

Does this analysis succeed in getting rid of “cognitive” or “rational” features and in replacing them by purely societal determinants culled from (micro-)sociology or anthropology? First, we notice that Latour gives a very unsatisfactory picture of scientific debates at the laboratory level, reflecting a far too restricted notion of scientific rationality. The contingency and irrationality of the scientific process, as presented in the text, are largely products of the outdated epistemology that is used to frame that process. Latour and Woolgar are at pains to show that actual science does not proceed by “deduction” from data (*op. cit.*, p. 136); but then, nobody ever thought that it did – at least not since the days of Descartes and Hobbes.

As a matter of fact, Latour’s findings are not contrary to a rational micro-level analysis of science, if we substitute a more up-to-date epistemology of science. Indeed, the authors’ own records easily allow reinterpretation in terms of a Popperian rationalistic methodology; science is a matter of making daring and risky conjectures and then leaving it to friendly collegial criticism to eliminate the invalid ones. Latour would later, in a less polemical vein, describe science as a matter of constant “trials of weakness”; this is easily translated into the Popperian conception of science as a constant probing for the weaknesses of daring proposals.

Latour would surely retort that this misrepresents what goes on in scientific practice, at the most fundamental level. At the level of the individual scientist, the purpose of collegial criticism is not the furtherance of the lofty institutional goal of science – the discovery of Truth – but rather to gain individual credit. This is indeed what Latour and Woolgar claim to document on their anthropological field-trip. The glee with which scientists dispute other researchers' findings and undermine their hypotheses shows that the discrediting of a rival project, and the ensuing increased credit of one's own project, are the real objectives.

However, even this will not suffice to prove Latour's point. He needs to "reconstruct", in micro-sociological but a-rational terms, not only the general critical and adversarial *modus operandi* of science, but also the particular forms that such criticism and evaluation take within specific "paradigms". Together, these elements make up what Kuhn calls the "disciplinary matrix" that is unique to each paradigm-based scientific community. Among the elements of the matrix are certain methodological "values" defining methodological propriety and correct procedure within the community. These are essentially macro-features of the research community in question, since the individual researcher is not at liberty to articulate his or her own version of these values; they are by necessity treated as being above individual preference and are imposed upon each researcher, however implicitly.

Moreover, there is a need for including the very theories that are invoked at the research frontier of the discipline in question, which also belong to the "disciplinary matrix". It is impossible to explain the professional interactions between the leading researchers in any field without a grasp of the theoretical basis shared by them, since it defines the problem context within the debates that are played out, including identification of the "anomalies" that are tackled within the framework of ongoing "normal science". Without an understanding of these elements, the very point of the researchers' endeavours will be incomprehensible, since it is essentially couched in terms derived from the theories. This is amply illustrated by Latour's own case study. The goal of the efforts he describes in such loving detail may be defined as "identifying the chemical composition of Thyropin Releasing Factor", which is incomprehensible lingo to anyone not familiar with such terms and the theories that are "black-boxed" in them. This theoretical background defines which moves are permissible and which impermissible in research practice, and which arguments are inconclusive and which carry the day. Hence, it is imperative to explain the fine-structure of the procedures and arguments in the laboratory that Latour documents with such care.

Popperians would locate such theoretical items in a Third World of objective problems situations, to avoid a psychologistic interpretation of science, but we should eschew such ontological profligacy and instead locate them at a societal macro-level: The theories are constituted by the commitment of the research community to certain semantic, logical and model-theoretical structures. The ontological status of the latter may be moot; but it is at least clear that the overall societal phenomenon is not located at the anthropological micro-level at which Latour moves in his investigation of laboratories.

Thus, the fact that Latour's rendition of the scientific process does not show us the impact of rational arguments, whether general or paradigm-specific, does not prove that belief in such arguments is mythical, or belongs merely to the rhetoric of science. Its absence is rather an artifact of Latour's own rendition of the scientific process, which is framed by an outdated epistemology directing us to look for deductions from facts to theories. Such will clearly not be forthcoming; but then, all philosophy of science of the last couple of generations would prepare us for their absence.

This constitutes a double failure for Latour. First, he has not achieved an explanation of science in *general* social terms – more precisely, micro-sociological or anthropological terms – which was the specific objective of STS. Hence, he has not achieved the goal that is shared by all representatives of science studies: to demonstrate that science is unexceptional and just one social institution among others. Neither has he achieved that integration between the different levels of societal description and explanation that is the goal of methodological individualism. Rationality and the specific theoretical commitments in a scientific community that define their governing paradigm belong to a different level, even if it is conceded that such items are not eternal, Platonic entities, but are socially and historically variable.

## 11.5 Enter the Actants

Perhaps as the result of these problems, the modest methodological step down the micro-macro ladder inherent in Latour's anthropological turn would soon be overtaken by a much more radical development: the introduction of the notion of *actants*. This was a loan from his collaborator Michel Callon, who had used it in a celebrated study of scallop fishing in St. Brieuc Bay in Normandy (Callon 1986).

Originally, however, the introduction of actants was intended by Latour to solve another problem in STS: the role of *nature* in the growth of science. As Latour read the Strong Programme, it had denied nature any role (this reading is dubious, however, but we need not here go into the issue of the correctness of Latour's reading). Against this, he urged that STS adopt a *symmetrical* stance towards nature and society as sources of input to scientific theories, akin to the recommendable Symmetry Principle of the Strong Programme that dictates symmetrical treatment of accepted and rejected theories in the history of science (Latour 1987).

The reintroduction of nature as a factor in the genesis of theories is controversial in the context of STS methodology, however. In the first place, the admission that nature plays a role in theory generation in natural science means giving up the rhetorically powerful slogan that theories are mere "social constructions". This is not so significant, however, since reservations on this point were present from the start in STS writings (Bloor 1976/1991). Next, a more serious worry: STS followers would have to give up some of their disciplinary autonomy, as they would have to rely on the expertise of natural scientists for a specification of those inputs to the

scientific process that come from nature. Such input typically, in advanced science, comes from sophisticated experiments, so STS accounts would have to contain detailed descriptions of experimental setups, the research work conducted by means of them, and the theoretical background that makes those experimental findings relevant. STS accounts would end up looking like standard reports of scientific research, which would tend to undermine the methodological distinctiveness of STS and cede crucial academic turf to the natural sciences.

This point was raised by Harry Collins and Steven Yearley in the so-called “Chicken” Debate (Collins and Yearley 1992a, b; Callon and Latour 1992), which was basically a confrontation between Strong Programme orthodoxy and Latourian heresy. As Latour and Callon would argue in this debate, the actant view is perfectly suited to handle the worry that STS would end up deferring to natural scientists: It will allow STS followers to include natural entities among the items causally involved in the genesis of natural science theories, yet without relying on natural science for the theoretical specification of those items. Instead, the theoretical description comes from a highly abstract theory that is social rather than physical; or, better, the very distinction between the two is dissolved. For actants *per se* belong neither to the physical nor the social realm.

Once this solution is in place, it could easily be deployed to solve the problem of rationality also; indeed, to Latour, the two problems were really two sides of the same coin. First, the problem of how to include items in the physical world without compromising the principles of STS is solved by describing the former in a terminology that treats them as quasi-agents, or *actants*, thus taking them out of the hands of the scientists and making them a bona fide part of (a reformed) social science. Second, traditional human agents are recast as actants, thereby subjecting them to ontological demotion since their rationality and even their intentionality are now stripped away; these too are just revealed as actor-network effects.

The notion of actants is imported from literary theory and many readers of Latour thought he had now retreated into the dark thickets of semiotics (cf. Pickering 1995, p. 12 ff). This interpretation was natural since, in a subsequent work presenting an extensive historical case study demonstrating his methodology in practice – an account of the conquest of France by Pasteur’s microbiology – Latour would actually describe his approach as “semiotic” (Latour 1988a, p. 11). However, in the second part of the same work, Latour made it clear that something far more ambitious was afoot. What was offered was rather a fundamental meta-physical solution, at the level of philosophy. Significantly, this part of the book was named *Irreductions*.

## 11.6 The Metaphysics of Actants

Latour’s category of *actants* is so abstract as to subsume virtually any entity, or kind of entity. As their name hints, actants are characterized by being active; their constitutive feature, and indeed their sole such feature, is that of *acting* upon

other actants. As Latour makes clear (or would later make clear), however, this to him is basically the equivalent of saying that actants are existent items, since to exist (to Latour) basically means having an impact upon other things.

The all-encompassing nature and hence near-vacuity of the actant category suggests that the point Latour is concerned to make is different from that of traditional philosophical ontology, where careful differentiation between – and classification of – various types of things are the goal. To Latour, actants are not a specific *kind* of thing, to be distinguished from other such kinds, but are rather the total set of the *particular entities* making up the world. More precisely, they are such particulars as viewed in their interactions with other particulars. Actants are what analytic philosophers would call “bare” particulars: items that we refer to simply as so many distinct particulars but without attributing to them any further properties as belonging to them *essentially* or *inherently*, *qua* actants. Actants obtain whatever further properties they possess through their interactions with other actants, which makes Latour’s ontology radically relational and dynamic, in contrast to the substance-based and static character of traditional philosophical (for example, Aristotelian) ontology.

Latour rejects potentialities and possibilities (*Irreductions*, §§ 1.5.1, 1.5.1.1), thus making his ontology radically actualist. To the potentialities belong logical powers, since drawing logical inferences means extracting – “de-ducting” – what lies hidden in the premises, implicitly and potentially. Consequently, Latour denies the reality of deduction (*ibid.*, 2.1.2). From this follows the fictitiousness of the epistemic processes that make up the “rationality” of traditional philosophy of science, for these are thought of as “deductions” or at least “inferences” from premises to conclusions. It follows that there is no knowledge, in the traditional sense of the term; nor are there theories (§ 2.1.7), or any explanation of individual cases that would be delivered by such theories.

The non-existence of knowledge and theory flows from another source as well: There is no representation, since it would constitute a meta-level with respect to the reality known (§§ 1.1.5.3, 1.1.5.4). But there are no such meta-levels, Latour’s ontology is perfectly flat. As there are no potentialities residing “inside” things, there is nothing “higher” than actual reality, either – no transcendent principles governing it and no meta-language in which to express such principles. Nor is there anything deeper, such as “underlying ontological structures”, or other foundations (*ibid.*, §§ 2.1.7.2, 2.1.7.3).

Latour rejects essences or universals (§ 1.4.2), accepting only particulars and unique events: “Everything happens only once” (§ 4.4.4). This goes for meanings, too (§ 2.6.3). Latour’s philosophy is thus radically particularist and actualist.

Latour’s actant ontology allows him to do away with the distinctions between intentional human subjects and objective natural items as an essential and inherent one. The features that render humans special – such as intention and thought – are themselves “networking effects”, the result of the association of actants; indeed, all properties of actants are networking effects (apart, we must assume, from their basic tendency to form networks). The actants are all of one type; whatever seems to set one kind essentially apart from another – such as intentionality, or a specific



mental capacity – is itself the result of the interaction between actants (Latour 2005, p. 204 ff). Thus, Latour would reject any attempt to capture what the actants are intrinsically, as opposed to what they do. That distinction is constitutive of a substance metaphysics that Latour is concerned to reject.

The points just made explain why Latour may justly call himself an *irreductivist*. Actants do not make up a uniform, abstract *kind*, rendering them all similar; they are just the sum total of *particulars*, and as such may be totally dissimilar in their concrete features. Such differences, and the richness and multiplicity they bring with them, must be respected by a philosophical inventory of the world; metaphysics must shun reduction. True, “bare” actants are featureless, by themselves, apart from their tendency to interact with other actants; but this does not matter, since any actant is invariably involved in such interactions: a non-interacting actant would be a non-existing one.

To Latour, the elimination of rationality, deduction, evidence, meanings, and so on, represents more than just a step towards methodological austerity in social science. All these things are “parts or powers of the mind”, according to traditional philosophy; and to Latour, their elimination is part of a general philosophical campaign against the notions of subjectivity, mind, intentionality, and the like, and their dominating role in European thought. More precisely, it is a campaign against the *division* of the world into two spheres, that of subjects and that of material objects, and the assumption that this division is fundamental, essential, and inescapable. To see this division as fundamental and to try to enforce it intellectually, socially, and politically is the hallmark of Modernism, according to Latour, and this imposition, referred to by Latour as the Modern Constitution, has done tremendous damage to European thought. Its elimination is the main objective of a subsequent work from Latour, *We Have Never been Modern* from 1993 (original French edition 1991). The notion of Nature as the realm of mute things and of Society as the abode of Subjects emerge at the same time, both are abstractions and, as such, falsifications. From this division springs most of the ills of the Modern condition: Its endless epistemological worries, with the subject (the Cartesian ego) being forever locked up in the prison of its own ideas and forever worrying about their veridicality, or indeed about the reality of the external world and of other minds. This quandary reaches a high point in the Kantian system, which at the same time serves as a *reductio ad absurdum* of the entire conception, however, since the subject and object end up being so rarefied that they slip out of the accessible, phenomenal world entirely, in the process being transformed into a ghostlike noumenon and the unknowable “Ding an sich”, respectively .

The Modern Constitution also blocks an adequate understanding of social reality, which ends up exclusively on the “mental” side of the divide as a system of abstract normative relations between individuals and social institutions, but floating free of any material constraints. But social life is deeply embedded in the material world and would not be possible without it; indeed, the world is made up of hybrids that bridge the gap between Nature and Society forced upon us by the Moderns. Modernity consists in the denial of the existence of such hybrids – or, better, in the refusal to recognize that there is really no duality to be bridged by hybridization in the first place.

## 11.7 Reassembling the Social

Above, we have traced various elements of Latour's thought as they were gradually shaped through his struggles with fundamental problems in the sociology of science. They comprise metaphysical and methodological elements as well as purely sociological insights. All these elements come together in Latour's recent book, *Reassembling the Social* (Latour 2005). Here, we get a detailed picture of Latour's metaphysical stance when transformed into a concrete methodology for social science. (Latour has expressly expanded his scope to social science in general in this work, thus going beyond the sociology of science.) In particular, we get a presentation of the full range of methodological intuitions propelling the urge for a microscopic analysis, articulated with Latour's inimitable flair for inventive metaphor.

First, he dismisses the idea that society is a specific, *sui generis* entity, as distinct from the collective of agents and their mutual interactions: "[ANT] does not take for granted the basic tenet of [traditional sociology]. It claims that there is nothing specific to social order; that there is no social dimension of any sort, no "social context", no distinct domain of reality to which the label "social" or "society" could be attributed. . ." (*Op. cit.*, p. 5). In summary of this harangue, he even goes so far as to endorse, somewhat jokingly, Mrs Thatcher's famous statement that "there is no such thing as society" (*Ibid.*, p. 6). Latour sometimes puts this familiar critical intuition of methodological individualists as a rejection of the idea that the social is a particular kind of "stuff" (p. 115, p. 159).

Instead, society – or, as Latour prefers to call it, the *collective* (*Ibid.*, p. 14) – is the global network of entities that are not themselves social but are, as we have learned, *actants* of countless different kinds. Or, better still, society is the very *process of networking*, or of forming associations between actants (p. 64 ff); it is a matter of "assembling the social", as the title of the book says. This calls for no less than a redefinition of the very discipline of sociology (p. 2) as precisely the tracing of those associations that form the collective.

Talk about society, social structure, and so on, is still in order, however, as long as it is clearly understood to be shorthand for a more detailed account dealing only in concrete particulars (p. 11). Still, this way of talking is dangerous, according to Latour, because it is easily taken literally (p. 67 f.) and will tempt us to invoke society to explain social processes. But this is to put the cart before the horse, mistaking the *explanandum* for the *explanans* (p. 8). "Society" is always what needs to be explained, while having no explanatory power in itself.

Latour puts it thus: "The second position [that is, ANT] takes as the major puzzle to be solved what the first [that is, traditional sociology] takes as its solution, namely the existence of specific social ties revealing the hidden presence of some specific social forms. In the alternative view, "social" is not some glue that could fix everything including what the other glues cannot fix; it is *what* is glued together by many *other* types of connectors. Whereas sociologists (or socio-economists, socio-linguists, social psychologists, etc.) take social aggregates as the given that could

shed some light on residual aspects of economics, linguistics, psychology, management and so on, these other scholars, on the contrary, consider social aggregates as what has to be explained by the specific *associations* provided by economics, linguistics, psychology, law, management, etc.” (p. 5).

This leads on to a second methodological point. In thus standing the approach of traditional sociology on its head, ANT insists on reference to *concrete mechanisms* whenever sociological macro-theory invokes such things as “society”, “classes”, “power”, “structure”, or “norms” to explain individual social actions. We need to be shown the concrete associations through which these ontological abstractions are constituted and through which the collective impinges upon the individual actant. Latour’s sociological stance is governed by a principle of “No social action at a distance”, he dismisses as pure mystification talk of individual action being determined by “social forces” that are directed towards the individual from a mysterious “social context”, and insists that such talk be replaced by detailed accounts of the concrete mechanisms at work (p. 167 ff.). Every scientific account deals with the individual case; there is science only of the particular (p. 137, p. 216).

Latour embellishes this idea with a number of colourful metaphors, recommending that sociology take the slow and circuitous byways in documenting the connection between actants, instead of travelling quickly along intellectual highways defined by abstract “forces” and “structures”. He also describes the use of such speedy routes as intellectual slight-of-hand and as a refusal to put in the required amount of hard work and scholarship; he urges us to pay for all our results with hard intellectual cash.

In traditional social science, the anti-holistic argument is suddenly halted when we reach the level of the human agent. Latour pushes on, however, dissolving even the individual into a network of actants. This heretic move was prepared in Latour’s general metaphysics, as laid out in *Irreductions*, where he dissolved the philosophers’ beloved “subject” (or mind, *res cogitans*, or intentionality) and its capacities into a network of simpler actants.

On the methodological level, Latour’s critical attitude to individual subjects is entirely in line with his opposition to the traditional macro-sociological approach. Just as he objected to explanatory strategies that invoke Society as some mystical entity beyond individuals, or some mysterious “stuff” with exceptional causal powers, he also objects to mental explanations invoking the mysterious entity of the “soul” or “subject”, or the magical properties of “mind-stuff”, the medium of such mystical operations as intention, deduction, rational evaluation, or reference. Like explanations that appeal to the magical powers of the “social”, mental explanations come too cheaply. They do not tell us *how* the mind concretely and operationally achieves its cognitive feats.

In *Reassembling the Social*, Latour adds some detail to these abstract metaphysical and methodological points: “You don’t have to imagine a “wholesale” human being having intentionality, making rational calculations, feeling responsible for his sins, or agonizing over his mortal soul. Rather, you realize that to obtain “complete” human actors, you have to compose them out of many successive *layers*, each of which is empirically distinct from the next. Being a fully competent actor comes in

*pellets*, or, to borrow from cyberspace, *patches* or *applets*, whose precise origin can be “Googled” before they are downloaded and saved one by one.” (Latour 2005, p. 207).

This argument does not, however, lead Latour to commence a project of redescribing social reality in terms of his newly discovered entity, the actant; the long sought-after Higgs particle of social physics, as it were. As Latour never tires of stressing in *Reassembling the Social*, ANT does not involve reconstructing social reality out of a specific *kind* of building block, but is a specific methodology that traces the concrete (particular) connections between actants. We might refer to this as a radical *methodological particularism*. All actants belong to the same category only in a philosophical sense; that is, in being all *particulars*. But particulars are not a certain *kind of thing*; any kind of thing is precisely a *kind* – hence a universal and not a particular.

In this way, Latour can combine his strong micro-analytic stance with a radical and explicit *anti-reductionism*. He insists that we respect the richness, variety, and concrete detail of the social reality (the collective) that we are called on to explain when doing social science. He objects to the way that theoretical explanation in traditional sociology reduces and falsifies the reality with which it deals: “While other sciences keep adding causes to phenomena, sociology might be the only one whose causes risk having the strange effect of making the phenomena they are supposed to explain vanish altogether” (p. 100). This happens because, in traditional sociological explanation, one entity is *substituted* by another one (*Ibid.*). Typically, theoretical explanation in sociology replaces “mediators” with “intermediaries”, where the former are unique, particular factors (actants) that leave their individual imprint on the outcome, while the latter are mere conduits of general causal forces, adding nothing to the outcome (Latour 2005, p. 105). At most, such individual peculiarities are conceptualized as mere statistical “perturbations” surrounding the true theoretical values and thus having no scientific significance (p. 152). Such a strategy goes directly contrary to ANT methodology: A social actor that makes no difference to the outcome is no actor at all, according to ANT. Every actor is a unique item, totally irreducible to any other (p. 153).

## 11.8 Methodological Lessons from ANT

Above, I have given an overview of Latour’s work, with special emphasis on the micro-macro issue that is central to it. What methodological lessons, then, can be learned from this examination?

First, let us look at Latour’s most notorious invention, his introduction of a more inclusive notion of social agent, the *actant*, by which he dissolves the (metaphysical and methodological) distinction between human agents and items in the material world surrounding them – along with many other distinctions. Here, Latour attacks the traditional narrowness of methodological individualism that has served as a straitjacket from which it is still struggling to free itself. As is apparent also from

some contributions to the present volume, even the admission of relations between social agents is a step that some methodological individualist have been loath to take; and agents' relations with the physical objects that serve as props for and a backdrop to their social interactions is largely absent from standard definitions of methodological individualism. Such objects are only granted a place in social ontology to the extent that this is bestowed upon them by specific social practices (the general logic of which is analyzed in Brian Epstein's contribution to the present volume, under the name of "anchoring"). But this implies that a material object, such as a large plastic container, only constrains our social interactions to the extent that we grant it the social status of being e.g. a recycling bin (Epstein's example), and not when its physical presence on the sidewalk forces pedestrians into awkward manoeuvres to avoid bumping into it, or into each other. This little toy example must be multiplied a millionfold, of course, to capture the way that our social interactions are constrained (but also facilitated) by material infrastructure such as buildings, roads, bridges, railways etc. and the vehicles that run on them; not to mention rivers, swamps, mountain ranges and oceans. So radical has been this conceptual agnosia vis-à-vis material reality in traditional sociology that Latour could create a stir by publishing an article in which he pointed to the existence of *doors* as a crucial factor in shaping familiar patterns of everyday human interaction (Latour 1988b. Notice how Mark Risjord makes a cautious step in the same direction in his contribution to the present volume, under the title of the *affordances* that material things offer agents; even illustrating it with the Latour-like example of a window affording a way of escaping from a room). Actor Network Theory has long since moved beyond such simple toy examples, having inspired an immense and ever growing body of work in sociology, in particular studies of technology and its societal role.

A second methodological lesson: Latour shows that the human individual is indeed a fairly arbitrary stopping point of theoretical analysis, if you are motivated by the familiar metaphysical, or metaphysical-cum-methodological, concerns that I spelled out initially. Latour is correct, I believe, that many of the metaphysico-methodological scruples that make some social scientists eschew explanation by reference to "Society", or to the powers of the "social context", apply equally or even more to explanation of human action in terms of intentions, knowledge, inferences, rationality, and desires of the human "mind" as the source of such processes.

This similarity is clearly illustrated by the works of Daniel Dennett, a current champion of (soft) reduction in the human sciences. Dennett's reservations about intentional explanation of human actions are strikingly similar to Latour's worries about "social explanation", as traditionally conceived. According to Dennett, such explanations tend to be uninformative, since they fail to show us how an action is actually accomplished. How does the "mind" actually manage to "recall" a name that it has temporarily forgotten, "intuit" the solution to a difficult problem in mathematics, "realize" that the thing on the path in front of it is not a snake but a twig, or just to construct a simple sentence in its native language? These labels masquerade as explanations, but really are no such thing. To genuinely explain such

accomplishments, we need to reduce them to steps so elementary that a simple mechanical device could do them, or, in more up-to-date terms, a computer capable of carrying out only simple binary operations (Dennett 1978). This is quite similar to the way that Latour requires social explanation to reveal the way that actants concretely bring about an outcome, by collaborating within a network, while eschewing glib explanation in terms of such macro-items as “capitalism”, “Empire”, or “norms”. On this point, his stance shows similarities with the “localism” recently proposed by Little (2012a, b; cf Little’s contribution to Chap. 4 in this volume).

Latour’s criticism also has affinities with Jerry Fodor’s attempt to dissolve the orthodox notion of the mind as a unified entity with general, multipurpose powers of thought. Fodor suggests an alternative account depicting the mind rather as a loose cluster of “modules”, each performing a different cognitive function, perfected through evolutionary history, some of them subsequently to be integrated with other such modules, but many of them remaining fairly insulated (Fodor 1983). Such modules are modelled upon dedicated, single-purpose computer programmes or “applications” and show a clear likeness to Latour’s “applets” and “downloads”.

A similar position is suggested in the works of Mark Turner (2001). Turner is especially concerned to show that the traditional view of the actor as a rationally integrated and consistent being is an illusion, and that we should turn to cognitive science for a more truthful, if perhaps less cohesive, picture of the human agent (cf. McCubbins and Turner’s contribution to Chap. 13 in this present volume).

Thus Latour’s work directs our attention to a structural weakness in traditional efforts towards micro-analysis within the sciences of man, broadly conceived to cover both the humanities and the social sciences. These efforts have moved within two different and mutually insulated registers, one in which social reality is reduced to human reality, the other in which human reality is reduced to something sub-human, whether biological, physiological or information-theoretical. What the former sees as the uncontroversial stopping point of analysis appears eminently deserving of further analysis, according to the latter. Latour is one of the very few thinkers who have attempted to combine the two registers – and in a particularly radical manner.

The conclusion is not, I believe, that social science is compelled to take its reductive efforts further, but only that it must look more pragmatically upon the notion of “methodological individualism” – a conclusion similar in spirit to that reached by other contributors to the present volume, although by different routes (cf. the Chaps. 9, 4, and 10 by van Bouwel, Little, and Zahle). Making the human individual the basic notion of social science is not a position forced upon us by metaphysical and methodological principles, but is rather pragmatically motivated. Or rather, while there may be good, principled reasons for explaining events at the macro-social level by reference to items at a lower level – reasons reflected in the anti-holistic intuitions mentioned initially – there is no principle dictating which lower level to choose, or how many steps to take down the ontological ladder. This is basically a matter of division of academic labour.

Here is a further methodological lesson: Latour's writings serve as a warning against a tendency in social science to reify a hierarchical or "stratified" view of social reality; as if social reality comes along in neatly pre-defined layers, ready for the social scientist to dig into, like an archaeologist stepping into a find site. Such a picture may be approximately true as long as we stay within a single discipline, such as economics, but it hardly makes sense when we take a look across social science as a whole. For instance, how on a putative single scale from "individual" to "whole" should we place the economists' macro-economic phenomena *relative* to the "Society" of holistic sociology, or the "Political System" of political theory? There is no unambiguous answer to such questions. According to Latour, all these reified social "strata" are to be considered as different but partly overlapping networks of actants, but without forming a neat overall hierarchy of levels; thus, the ontology of the social world is perfectly flat. The idea of replacing the hierarchical ontology of traditional social theory with a flat ontology composed of networks of different sizes is one that is currently being suggested by other philosophers of social science (Ylikoski 2012; cf. Ylikoski's contribution to Chap. 7 in this volume). But ANT is the only version that has been applied on a large scale in ongoing social research.

## 11.9 Troubles for ANT

Finally, it is inevitable that we ask whether or not Latour's radical proposal takes care of all the methodological individualist's worries? Is ANT what the methodological individualists dreamed of all along, only sharpened and deepened by Latour's more consistent observance of the principles of methodological individualism?

Unfortunately, I believe that the answer is "no". What Latour gives us in *Reassembling the Social* is not a workable recipe for a social science methodology (at least not yet). There are too many gaps and uncertainties, too much metaphysics, and too many apparent paradoxes. I will just mention a few points, below.

Latour's thought is driven by two strong metaphysical impulses, which unfortunately push him in opposite directions. One, the analytic impulse, leads him to postulate an extremely abstract and featureless entity, the actant, as the basic constituent of his system. Concrete features and properties only accrue to actants through their interaction with other actants. This picture is reminiscent of, and partly inspired by, Alfred Whitehead's ontology, as Latour himself indicates, and the entire approach recapitulates discussions in British philosophy of a 100 years ago between proponents and opponents of the view that things possess properties only through their relations to other things (with Francis Bradley and John McTaggart representing the former position, Bertrand Russell the latter). This stance is counteracted by an equally strong anti-reductivist conviction, however. Latour insists that everything is what it is and not another thing, to echo Bishop

Butler's famous slogan, and must be accommodated by science in its full individuality. Nothing may be reduced to or replaced by another thing.

The attempt to combine these two stances gives Latour grave problems, both at the metaphysical and the methodological level. Let us leave the philosophical issues on one side here and look only at methodology (for the metaphysical problems, see Harman 2009; Collin 2011). Latour's efforts to solve the methodological problems result in what we may fairly describe as a radical version of ethnographic method, adopting a purely "emic" stance. Agents' self-descriptions are to be respected, and Latour heaps scorn upon Critical Theorists in particular for their contemptuous dismissal of social agents' own conceptions as mere "false consciousness". But, although the stance adopted by Horkheimer and Adorno and their colleagues at the *Institut für Sozialforschung* was surely excessive, Latour's recommendation seems to err just as badly on the opposite side. Social science may surely at times have good theoretical reasons for disregarding agents' own reports. We recall that, in his early empirical studies, Latour would adopt an "irreverent" approach to scientists' own accounts of their activities, refusing to defer to them as authoritative (*op. cit.*, p. 29). And there is little doubt that empirical science studies have generally been salutary in highlighting certain aspects of the scientific process overlooked or suppressed in the official, idealized self-conception of science. Latour's methodological stance in *Reassembling the Social* seems a retrograde step compared to the one adopted in *Laboratory Life*.

Latour's problems would be partly alleviated if he had heeded certain distinctions familiar from the debates about micro-macro analysis and reduction in analytic philosophy of science. Following standard terminology, which labels all types of analyses as "reductions", the crucial distinction is that between *eliminative* and *non-eliminative* reductions. Most cases of reduction in natural science are of the latter kind, as when "heat" of a gas is reduced to mean kinetic energy of the gas molecules, or water reduced to H<sub>2</sub>O. There is no suggestion here, of course, that heat and water do not really exist.

There are actually two points being made here. First, empirical science as such is neutral on the metaphysical issue; it only tells us that we can, for instance, explain everything that is explained with reference to *water* by reference to H<sub>2</sub>O instead. The two explanations are connected by "bridge principles" stating that heat in a gas is identical with mean kinetic energy of the gas molecules, or that water is identical with H<sub>2</sub>O. Such identities do not imply that one of the items identified does not really exist; indeed, rather the opposite. To eliminate an item will call for a special argument, of a philosophical nature; thus, it is left to philosophy to draw the final metaphysical conclusions. This leads to the second point. In cases like the ones mentioned, philosophy will surely not declare such things as heat or water to be non-existent. Admittedly, this leaves philosophers with a thorny problem concerning the nature of the relation between the observational properties defining heat or water in an everyday context and their theoretically specified properties. To capture it, notions such as "emergence" and "supervenience" have been introduced; these are notions that are generally conceded to be in need of further analysis (cf. the Chaps. 2 and 7 by Epstein and Ylikoski in this volume). But most



philosophers resist the conclusion that these difficulties are so grave as to render non-eliminative reduction meaningless.

The problem that worries Latour is really quite a different one. Social scientists themselves – *not* philosophers of social science – will occasionally want to reject agents' reports as illusive and as providing distorted views of social reality. This is not the result of general metaphysical scruples, but merely illustrates that during the historical development of any scientific discipline, researchers will sometimes be justified in dismissing certain observations as non-veridical, for reasons flowing from the body of theory accumulating within that discipline itself. (Consider the way that astronomers gradually learned to disregard certain features of celestial observations as due to atmospheric distortion, diffraction, and similar effects.) Social science is no exception to this general rule. But Latour cannot extract an argument against this practice from his general anti-reductivist metaphysics.

Latour's predicament is thus aggravated by his failure to heed certain standard distinctions and methodological commonplaces from analytic philosophy of science. At a deeper level, it springs from his conflation of issues belonging to empirical science and to philosophy. But we must note that, to Latour, these are not oversights or confusions, but daring attempts to dissolve standard categories; he proudly presents his system as "experimental metaphysics" (Latour 2004, p. 123) which according to standard conceptions is close to an oxymoron. Let us end by noting that Latour's struggles with these problems is very much a case of work in progress; we have certainly not heard Latour's final word on these issues. In the meantime, we should appreciate the way that Latour's work serves as a laboratory in which novel methodological ideas, some of which are currently being suggested also by other authors, are being tested out on a body of concrete empirical material.

## References

- Bloor, D. (1976/1991). *Knowledge and social imagery* (2nd ed.). London/Chicago: Routledge & Kegan Paul/University of Chicago Press.
- Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St. Brieuc Bay. In J. Law (Ed.), *Power, action and belief: A new sociology of knowledge* (pp. 196–233). London: Routledge & Kegan Paul.
- Callon, M., & Latour, B. (1992). Don't throw the baby out with the bath school. In A. Pickering (Ed.), *Science as practice and culture* (pp. 343–368). Chicago: University of Chicago Press.
- Collin, F. (2011). *Science studies as naturalized philosophy*. Dordrecht: Springer.
- Collins, H. M., & Yearley, S. (1992a). Epistemological chicken. In A. Pickering (Ed.), *Science as practice and culture* (pp. 301–326). Chicago: University of Chicago Press.
- Collins, H. M., & Yearley, S. (1992b). Journey into space. In A. Pickering (Ed.), *Science as practice and culture* (pp. 369–389). Chicago: University of Chicago Press.
- Dennett, D. (1978). *Brainstorms. Philosophical essays on mind and psychology*. Montgomery: Bradford Books.
- Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- Harman, G. (2009). *Prince of networks: Bruno Latour and metaphysics*. Melbourne: Re.press.
- Latour, B. (1987). *Science in action*. Cambridge, MA: Harvard University Press.

- Latour, B. (1988a). *The pasteurization of France*. Cambridge, MA: Harvard University Press.
- Latour, B. (1988b). Mixing humans and nonhumans together: The sociology of a door-closer. *Social Problems*, 35(3, Special issue: The sociology of science and technology), 298–310.
- Latour, B. (1993). *We have never been modern*. New York: Harvester Wheatsheaf.
- Latour, B. (2004). *Politics of nature*. Cambridge, MA: Harvard University Press.
- Latour, B. (2005). *Reassembling the social*. Oxford: Oxford University Press.
- Latour, B., & Woolgar, S. (1979/1986). *Laboratory life*. Princeton: Princeton University Press.
- Little, D. (2012a). Analytical sociology and the rest of sociology. *Sociologica* 1.
- Little, D. (2012b). Explanatory autonomy and Coleman’s boat. *Theoria*, 27/2(74):137–152.
- Pickering, A. (1995). *The mangle of practice*. Chicago: University of Chicago Press.
- Turner, M. (2001). *Cognitive dimensions of social science*. Oxford: Oxford University Press.
- Ylikoski, P. (2012). Micro, macro, and mechanisms. In H. Kincaid (Ed.), *The Oxford handbook of philosophy of the social sciences* (pp. 21–45). Oxford: Oxford University Press.

## Chapter 12

# Structure, Agency, and Improvisation

Mark Risjord

**Abstract** The improvised joint performances of small-group jazz ensembles, basketball squads, or partners in conversation pose interesting challenges for conceptualizing the relationship between agents and social structures. Contemporary approaches to the structure and agency debate can be divided into microfoundational and practice-theoretic. Heir to the individualist tradition, microfoundational approaches treat structures as continually re-created by the instrumentally rational choices of agents. Heir to the holist tradition, practice theory treats social structures as reproduced practices and agents as having dispositions to reproduce those patterns (e.g. Bourdieu's habitus or Giddens' practical consciousness). In this essay, a detailed discussion of jazz improvisation is the basis for a critique of both microfoundations and practice theory. Neither way of conceptualizing structure and agency can account for the way jazz musicians maintain improvised musical forms. This essay suggests an alternative conception of agency as ecological attunement. In order to maintain structures improvisationally, agents must be capable of a kind of mutual coordination which is not the product of planning or deliberate choice. The essay proposes a three-part analysis of the capacity for mutual coordination. The resulting conception of agency shares features of both dominant approaches. Like microfoundations, it treats social structures as continually created by agents, not as reproduced patterns. Like practice theory, it refrains from conceptualizing agents as deliberate choosers. Ultimately, to be an agent requires treating others as agents and responding to the joint possibilities for action provided by the environment.

---

M. Risjord (✉)

Department of Philosophy, Emory College of Arts and Sciences, Emory University, Atlanta, GA, USA

e-mail: [mark.risjord@emory.edu](mailto:mark.risjord@emory.edu)

## 12.1 Introduction: The Dynamic of Structure and Agency

Jazz improvisation, especially as popularly imagined, represents a quintessentially individualist activity. The soloist stands before the band and creates something entirely new. A well-known textbook on the history of jazz, for example, characterizes it this way:

Jazz, like the blues, emphasizes individualism. The performer is at the same time the composer, shaping the music in style and form. A traditional melody or harmonic framework may serve as the takeoff point for improvisation, but it is the personality of the player and the way he or she improvises that produces the music. (Southern 1997)

Such a view of improvisation conflicts with the results of ethnomusicological investigations of jazz. Works like Monson's *Saying Something* (1996) and Berliner's *Thinking in Jazz* (1994) have emphasized the conversational and interactive character of ensemble playing. Through their joint improvisations, the jazz musicians produce and maintain a musical structure, of which the improvisation of a melody is only one, albeit prominent, aspect.

These competing pictures of jazz performance are a microcosm of the structure and agency debate within the social sciences. A number of social scientific programs treat social structures as the product of action, and take structures to influence the formation of agents. The social scientific problem is to unpack the dynamic relationship between structures and agents. How are social structures (including social roles and norms, enduring patterns of social interaction, as well as institutions) created, reproduced, or subverted by individual actions? And how are individuals formed or influenced by social structures? Philosophically, the structure and agency problem spawns conceptual questions about the key ideas: What are "structures" such that they can be created, maintained, and changed by agents? And what are "agents" such that they can both be formed by and be formative of structures?

The older dichotomy between individualism and holism is discernible in contemporary approaches to the structure and agency problematic. Heir to individualism is a group of approaches Daniel Little calls "microfoundational" (Little 1998, 1989). These views recognize that each human is born into a group of people who exhibit patterns of belief and activity. Contact with others thus facilitates the acquisition of beliefs, values, abilities, and attitudes. The choices open to an individual are partly set by the social institutions, roles, practices, etc. In these ways, features of agency depend on the social environment. Little argues that explanation in the social sciences is causal, and all social causal mechanisms "supervene upon the structured choices and behavior of individuals" (Little 2007, 360). Analytical sociology and public choice theory are clear examples of Little's microfoundational approaches. They look for explanations that articulate individual-level mechanisms for social processes, especially rational choice or game-theoretic mechanisms. On the other side of the debate, Pierre Bourdieu's practice theory and Anthony Giddens' structuration theory are heirs to structural-functionalist views in the social sciences. Like the microfoundational approaches,

these views treat social structures (practices) as patterns of activity, and they regard agency as always situated within a field of existing practices. In contrast to microfoundational approaches, however, they eschew the explanatory strategies of micro-economics, and do not directly appeal to causality in their explanations. The various forms of practice theory<sup>1</sup> are interested in the way that agents strategically use social structures in their interactions, and how they thereby reproduce, modify, or undermine the practices which constitute the social environment.

The philosophical differences between microfoundational approaches and practice theory turn on the conceptualization of both social structure and agency. Microfoundational approaches treat social structures as either the intentional or unintentional product of choices. The explicit agreement to form a club with officers, roles, and responsibilities would be an example of a social structure created by intentional choice. Alternatively, social structures could result as the unintended side effects of individual actions and choices. Thomas Schelling's (1969) account of neighborhood segregation as the unintended result of preferring similar neighbors would be an example. Practice theorists recognize that patterns of social interaction could arise in both of these ways, but as we will see below, they reject the underlying conception of agency. As a result, practices are supposed to be neither a deliberate creation nor an unintended side effect of other deliberate choices, but a recurring pattern of interaction. The intentionality of the action is not as important as the outward, molar form of behavior that agents represent and interpret for their own purposes: "regularized behavior reflexively sustained as such by its participants" as Giddens expresses it (1984, 14). The deep difference between the two conceptions of social structure is that the practice theorists treat practices as reproduced regularities while the microfoundational approaches treat them as consequences of underlying choices. Structures are not reproduced, according to the microfoundational views, but recreated anew by action.

The microfoundational and practice theoretic approaches conceptualize agency in fundamentally different ways. Microfoundational approaches see agents as choosers. The practical syllogism is taken as the paradigm of intentional action, even if some theorists feel the need to add emotion or other non-rational elements to their explanations (e.g. Elster 2007). By retaining instrumental rationality as central to the concept of agency, these programs can use the resources of economic models, decision theory, and game theory for micro-explanatory purposes. According to practice theorists, agents reproduce practices through their actions, but not typically through their *choices*. Practice theorists try to understand instrumental rationality as only one moment or mode of agency. Determining just what (else) agency amounts to has been an important theoretical problem for practice theory. Indeed, articulating a concept of agency may be one of the most important philosophical challenges of contemporary practice theory. Bourdieu and Giddens have attempted to

---

<sup>1</sup> For the sake of terminological simplicity, I assimilate Giddens' structuration theory within the scope of the phrase "practice theory." For precedent in treating Giddens in this way, see (Karp 1986).

reconceptualize agency in somewhat different ways. In Sect. 12.3, below, we will examine their views along with the conception of agency implicit in microfoundational accounts. As we will see, all encounter significant difficulties.

Bourdieu often returns to the image of improvisation to simultaneously express the freedom of agents within a structure and the way in which structures arise from free action. Bourdieu uses the metaphor to express the tension between conservatism and creativity which inheres in the conception of agency. Improvisation continuously invents new forms and, at the same time, preserves an established framework. In jazz, the musical structures are created, maintained, and transformed by the ongoing improvisation of the musicians. Unfortunately, Bourdieu never really embraces the power of the metaphor. Rather than speaking of improvisation simpliciter, he uses near-oxymoronic phrases like “intentionless invention of regulated improvisation” (Bourdieu 1977, 79) or “orchestrated improvisation of common dispositions” (Bourdieu 1977, 17). Apparently, the spontaneous character of improvisation needs discipline. In spite of Bourdieu’s hesitation, the musical metaphors of improvisation and virtuosity are generative. They portray agency as automatic, unconscious, and embodied, yet at the same time interactive and creative. Perhaps the way to develop Bourdieu’s metaphor is to look carefully at the phenomenon of musical improvisation.

## 12.2 Improvisation in Jazz

In common speech, jazz improvisation refers to the extended instrumental soloing so distinctive of small-group jazz genres. It is crucial to recognize, however, that the members of a jazz ensemble who are not soloing do not read their parts from written scores in the way that classical musicians do. All members of the ensemble are improvising their parts. This makes the activities of the “rhythm section”—typically the drums, bass, and piano—particularly interesting from the perspective of the structure and agency problem. Ingrid Monson’s *Saying Something: Jazz Improvisation and Interaction* (1996) focuses on the way in which the rhythm section establishes and maintains a musical structure on which the solo depends. Her work, therefore, potentially holds some clues to the analysis of improvisation that will shed light on the structure and agency problematic.

In small-group jazz ensembles of the late twentieth century, a typical song structure is to play the written melody of the song once at the beginning, and then once again at the end. The middle of the performance contains improvised melodies by the various members of the ensemble. While one instrument solos, the rhythm section performs two essential functions. First, they maintain a musical structure that provides the harmonic and rhythmic background for the solos. This musical structure is a sequence of chords, a cycle or chord progression, and it is often closely related to the harmonization of the original melody. In a standard quartet, the bass and piano are responsible for maintaining the sequence of chords in proper order. The drums support the structure by playing rhythmic variations that

anticipate and mark important transitions in the chord progression. Solos key to this harmonic and rhythmic structure by (typically) beginning and ending with the chord cycle. Second, the rhythm section maintains the tempo, and more importantly, the “feel” or “groove” of the performance. Groove is not easily articulated in words, but it is immediately recognized by the non-musician as what makes jazz, rock, or blues rhythmically distinctive genres. Both the harmonic structure and the rhythmic feel are continuously maintained by the actions of the individual musicians. Their actions are improvised in the literal sense that the vast majority are not planned in advance. Nor are they chosen as the means to an end. Hence they are not deliberate or purposive actions in this sense. Monson’s analysis aims to understand how these musical structures are created, maintained, and developed by the ongoing activity of the rhythm section.

One way in which social scientists, and especially ethnographers, have tried to study structure and agency is through looking carefully at breakdowns or ruptures in the social fabric. Many have argued that the dynamics of social forms are most visible in times of crisis, as agents work to reestablish or change structures. Monson provides an extended and detailed analysis of just such a breakdown and recovery of structure. Her analysis focuses on the recorded performance of “Bass-ment Blues” by the Jaki Byard Quartet in 1965. The melody for this performance is played by the bass, and is followed by a bass solo. The bass is a key element of the rhythm section, providing the rhythmic pulse for swing. When a bass carries the melodic line during a bass solo, the piano and drums must play a supporting role. Joint action of the musicians, soloist as well as rhythm section, maintains the harmonic and rhythmic structure.

During the performance analyzed by Monson, the musicians became out of phase: the bass player “articulates the harmonic structure of a twelve-bar blues . . . six beats ahead of where the original top of the chorus would have been” (Monson 1996, 159). The harmonic cycle is constituted by the notes that each player contributes to the whole sound, including the soloist. The location of the “top of the chorus” depends, then, on the joint action of the performers. The fact that the band is six beats out of phase is significant, because it means that the musicians had two problems to solve. In this song, each measure had four beats. A six beat difference means that the musicians were disagreeing about both where the chord progression began and about which beat is the first beat of the measure. The drums and piano were maintaining one version of the structure, and the bass another.

From the performer’s point of view the mistake was a potential disaster. Unless the musicians could recover, there would be a potential clash between the melody tones played by the bass and the chordal harmonies played by the piano. There would also be a critical ambiguity about the timing of the next solo entrance. The challenge for each player was to respond to the others in a way that re-established their unity, rather than either falling into an incoherent collage of individual action or stopping the music entirely. Over the course of four chord cycles the musicians re-established the harmonic structure. Not only did the musicians recover, Monson says that the mistake was invisible to both the audience and those who later produced the recording. She identified it only when she tried to transcribe the

recording. The musicians had sufficient virtuosity with the practice that they fixed the broken pattern without losing the musicality of the performance.

In describing the musicians' recovery, Monson notes several events. The disruption occurs during the third time through the cycle (or third "chorus") when the bass added six extra beats to the chord structure. The piano player then snaps his fingers at the point where, according to how he was keeping count, the beginning of the fourth chorus should be. At the beginning of the fifth chorus (again, according to the way the piano player was counting), the piano player changed his style of accompaniment. He began playing a single bass note on the first and third beats, and a full chord on the second and fourth. The drummer responded by simplifying his rhythmic accompaniment, emphasizing the bass drum on beats one and three. (This would sound something like the "boom-chick" of a two-step dance rhythm.) The simplification of the harmonic and rhythm patterns, however, was not sufficient to bring the band back together. Describing the final phase of their recovery, Monson writes:

The adjustment occurs when Tucker [the bass player] plays an obvious two-measure turnaround (3625; mm. 71–72) after completing the twelve-measure chorus that began in measure 59. Byard and Dawson [the piano player and drummer] respond to the musical signal by adjusting to Tucker: they add two beats to the twelfth measure of their sixth chorus, and the three musicians establish the top of the seventh chorus together (m. 73). (Monson 1996, 170)

A "turnaround" is a sequence of chords that harmonically resolve to the first chord of the progression. Jazz musicians immediately recognize a number of these standard forms, and the piano player and drummer would certainly realize that the bass player had begun to play a turnaround.

In the quotation above, Monson talks about a "musical signal." One might think that the best way to explain how the musicians re-established the musical structure is through signaling. There are at least three events in her description which could be interpreted as signals: the finger snap, the change in rhythmic feel, and the turnaround. There are several reasons why this is an unsatisfactory, or at best only partial, explanation. For, what would it be a signal of? What would its content be? The harmonic structure is not an objective form to be marked. The position of the beginning of the cycle depends on the ongoing interaction of the players. So, the finger snap cannot be interpreted as drawing the other musicians' attention to a fact. At most, it indicates difference: I am beginning here, while you are beginning there. The discrepancy, however, was almost certainly already common knowledge. So as a signal that there is a difference between the performers, it was redundant. Even if the finger-snap (etc.) did alert the others to the discrepancy (or, perhaps, indicated that the piano player was aware too), it leaves the crucial question unanswered. How would the signals help the musicians adjust their playing so that they re-established the structure? To do so, a signal would have to communicate intention in sufficient detail as to permit coordination; it would have to say "you do that while I do this." None of the possible signals have this level of complexity. To explain the recovery of the harmonic structure, more needs to be added.



Monson does not use signaling to explain the recovery. She interprets piano player's and drummer's change in rhythm as trying to "promote the reestablishment of the chorus structure" (Monson 1996, 170). She leaves unexplained how the chosen rhythm and harmonies would do so. I suggest, based on my own experience as a jazz musician, that the simplified rhythm and harmonies reduce ambiguity. Even though the sequence of chords is fixed, the rhythm section players have a wide range of rhythmic and harmonic choices. For example, the triad of a C major chord is the first, third, and fifth notes of the C scale: C, E, and G. When the chord progression specifies a C major chord, the piano player need not play any of these notes. Indeed, he or she will normally play other notes that stand in various harmonic relationships to the chord, making it sound simple and clear, or dissonant and complex. These alternate ways of playing the chord (different "chord voicings") create harmonic ambiguities in the sense that the soloist can play notes that would fit with many different chords. Ambiguous chordal accompaniment therefore opens the range of melodic possibilities for the soloist. Reducing ambiguity by playing simple chords reduces the range of melodic possibilities. Similar points hold for the drummer. Rhythms that stay away from the normally emphasized beats leave the soloist a larger space for syncopation. Returning to a march-like rhythm where the bass drum plays on the first and third beat with the snare on the second and fourth makes it more difficult for the soloist to emphasize other beats of the measure. The stylistic change by the piano player and drummer was thus a way of changing the possibilities of action by the soloist. They made it harder for him to stay out of phase, and thereby easier to get back together.

The change in style by the drummer and the piano player, then, promoted reestablishment of the structure by changing the possibilities of action for the soloing bass player. They changed the environment in which the bass player was acting. This explanation extends naturally to the two-bar turnaround. Turnarounds are not only recognizable, they are predictable. When the bass player starts the turnaround, the beginning of the cycle is guaranteed to begin (for the bass player at least) in eight beats. Therefore, while this may function as a signal, its meaning is not found in some information about the sender's intentions or instructions for the recipient. The significance is what it allows the other musicians to do. A two-measure turnaround gives the pianist and drummer eight beats of time to anticipate a new downbeat. Just as the change in accompaniment changed the environment for the bass, the turnaround changes the environment for the piano and drums. It makes possible the difficult task of adding two extra beats to a four-beat measure.

Monson's account of this musical breakdown and recovery is a rich resource for thinking about structure and agency. The structure is the ongoing joint musical performance, particularly the rhythmic and harmonic cycle of the chord progression. This is maintained by the action, and more importantly, the *interaction* of the agents. The agents produce, reproduce, and maintain the musical structure through their interaction. Interestingly, the interaction she describes is not a set of temporally sequential responses; it is not like a chess game. And while communication is part of the interaction, the maintenance of harmonic and rhythmic structure requires

more than signaling. How should we conceptualize the agents who are acting and interacting? How must agency be understood if we are to understand social structures as produced and maintained by action, as both the microfoundational and practice theoretic accounts do?

## 12.3 Agency Within Structure: Three Approaches

Any understanding of agency must come to grips with a fundamental tension. Social structures like implicit norms, ritual performances, games, even conversations and jazz performances, are relatively long-lasting patterns of action. And while no social structure is permanent, some remain stable over the course of several generations. Agency seems to need some characteristics that tend to reproduce or maintain these patterns of action. At the same time, agency is supposed to be a creative force that can challenge or change the established patterns. Each of the three conceptions of agency to be discussed here respond to the tension in slightly different ways.

### 12.3.1 *Agents as Rational Actors*

The microfoundational approaches aim to use the resources of rational choice theory and game theory to explain how agents create and reproduce structures. Contemporary approaches are not thin exercises with abstract models. The logic of decision is often supplemented with more realistic theorizing about the role of emotion or cognitive processing (Elster 2007; Bicchieri 2006). And as Little has argued, since existing social practices are an aspect of the environment that influences choice, microfoundational theories can appeal to social norms, values, symbolic meaning, and identity (Little 1998, 1989). Fundamentally, however, these views seek to understand the choices made by agents, and to show how social phenomena are the product of individual choices. It is this aspect of their understanding of agency that makes it unable to accommodate the interaction of jazz musicians during improvisation.

Conceptualizing agency in terms of instrumental rationality requires us to understand the musicians' response to the breakdown, as well as their ongoing maintenance of the harmonic structure and rhythmic groove, in terms of their choices. Certainly there were several key moments of choice: the finger snap, the change in harmonic and rhythmic complexity, and the final six-beat time shift which put the group back on track. Developing an analogy with Grice's theory of conversational implicature (Grice 1989) would be one natural way to apply the rational actor model. The musicians would thus be interpreted as communicating by means of musical signals. However, as has already been argued, such an interpretation does very little to advance our understanding. Nothing in the finger snap,

change in rhythm, etc. can be interpreted as a signal about *how* to reestablish the form. At best, they are signals *that* there is a problem, which was already common knowledge.

Microfoundational approaches can give very sophisticated accounts of the way that choices interact to produce patterns of social action. However, neither maintaining the musical structure nor recovering from a breakdown can be modeled in game-theoretic terms. The payoffs must be linked to the wrong sort of outcomes. There may be some kind of strategic interaction of choices involved in deciding which song to play at what tempo, or what to wear to the gig. Here, one can imagine payoffs of various sorts. But it is not clear what the payoffs would be for voicing a chord in one way or another. Monson's account highlights the central importance of mutual responsiveness for the interaction. Each of the musicians is responsive to what the others are doing and is actively changing the environment of the other players. Through this ongoing coordination, they maintain the musical structure. This sort of interaction is invisible from the perspective of instrumental rationality. Therefore, even if we want to retain rational choice as a part of our explanatory apparatus, it needs to be substantially supplemented in a richer account of agency.

### 12.3.2 Bourdieu's Habitus

Bourdieu's conception of the "habitus" is at the heart of his conception of agency, as well as his response to the structure and agency problematic. Like the microfoundational approach, Bourdieu sees practices as created and reproduced by the actions of the agents. However, he does not think of those actions as always deliberate choices. While agents do use existing practices strategically, manipulating implicit and explicit norms for their own ends, they do not maintain the patterns through their choices. Bourdieu's notion of "habitus" can be interpreted as a "common code" (Bourdieu 1977, 80) or internal mechanism by which individual agents reproduce social practices. The habitus is a body of dispositions that serve as an "organizing principle" (Bourdieu 1977, 18), shaping thought, perception, and action. The habitus is

a system of lasting, transposable dispositions, which, integrating past experiences, functions at every moment as a *matrix of perceptions, appreciations, and actions* and makes possible the achievement of infinitely diversified tasks. . . (Bourdieu 1977, 82–3 emphasis in original)

While the details of individual history will vary and produce slightly different dispositions, the outward form must be similar enough to reproduce the pattern. The explanatory formula thus runs from patterns to dispositions and back to patterns. Each individual's dispositions are formed by the existing social structures. Individuals end up with similar dispositions, and these dispositions then recreate the very patterns of action which produced the dispositions in the first place.

Bourdieu's analytical framework has been applied to jazz, but the object of analysis has been at the level of cultural production: how jazz as an art form or performance genre is shaped by social factors (Lopes 2000; Banks 2012). Such analyses overshoot the mark. Monson's problem is to account for the maintenance of specific musical structures, not the characteristics of jazz music as a social practice. Nonetheless, Bourdieu's conception of habitus would seem to be tailor-made for improvisation. Each musician has, through long practice, developed a set of dispositions that are invoked by the performance. To maintain a rhythmic groove, for instance, each musician has to be sensitive to subtle differences in the timing of the notes played by the others, and be prepared to respond in a genre-appropriate way. We can thus explain how, for example, a swing feel is maintained through the performance, and how it is replicated from performance to performance.

Appealing to the musicians' habitus has a better chance of explaining the phenomenon of musical improvisation than appealing to their deliberate choices. There is an important sense in which the actions which make up a jazz performance are automatic, and the habitus expresses that sense. However, we have already noted the tension between creativity and conservatism which haunts any account of agency within structure. And where deliberate choice over-emphasizes the creativity of improvised action, the habitus over-emphasizes reproduction. When everything is flowing smoothly, it is natural to understand the musicians as being on autopilot, tuned into the harmonic and rhythmic features of the other musicians' performances and responding with their own harmonies and rhythms. The habitus is less informative when we try to account for the recovery from a rupture in the smooth flow. No doubt that there is some set of dispositions linking perception and action which underlie the musicians' capacity to recover. But such an appeal to dispositions merely invokes a powerful unexplained explainer. It is nothing more than a placeholder for a deeper explanation.

The root of the problem with habitus is twofold. First, Bourdieu's conceptualization of agency relies on the notion of a disposition. Dispositions are convenient conceptual tools for linking perception and action, but they are always explanatory placeholders. This is unproblematic when we do not care about the micro-level details. For Bourdieu's analysis, it does not matter how the habitus is implemented in cognitive-level or neural-level mechanisms. The important characteristics are the macro-level patterns of behavior, which are explained by the inputs and outputs, not the mechanisms which link them. In Bourdieu's framework, the inputs are experience with past practices and the outputs are actions which reproduce those practices. The notion of reproduction—the second root of our difficulties—requires something existing at one time to be re-created at another. This pair of commitments leads to familiar criticisms of Bourdieu that the habitus is nothing but social structures internalized (Turner 1994; King 2000). For our problem of understanding improvisation, the commitments lead to an explanatory impasse. While the group improvisation is going smoothly, it treats the agents as unthinking automata; when they are trying to recover from a breakdown, it provides no resources. The conception of agency which takes the habitus as its central construct is thus not

adequate to support the dynamic of structure and agency implicit in an improvising jazz combo.

### 12.3.3 *Giddens and the Levels of Consciousness*

Readers of Anthony Giddens' many works will recognize the problem with which we have been engaged as "structuration," that is, the process by which the social system at one time is reproduced at another. While Giddens' view shares much with Bourdieu, at least as I have interpreted him, I would argue that Giddens has done more than Bourdieu to show how the concept of agency needs to be radically transformed. Giddens critiques the rational actor model of agency, and argues that it misunderstands the relationship between agency and action. The rational actor model defines agents as those who act intentionally. Giddens suggests inverting this relationship by taking agency as primitive and defining action in its terms: actions are the "stream of actual or contemplated causal interventions of corporeal beings in the ongoing process of events-in-the-world" (Giddens 1979, 55). There is a striking parallel between Bourdieu's conception of habitus, as interpreted above, and the rational actor model of agency. Both treat actions as the result of internal processes—habitus for Bourdieu, beliefs and utilities for the rational actor model—and then treat agents as those beings who act. Giddens' prioritization of agency over action might permit us to sidestep the problems identified above. But what does it mean to think of actions as the agent's interventions in the flow of events? We have returned to our original question about how agency is to be conceptualized.

Giddens approaches the question of agency through a three-part analysis of consciousness (developed in Giddens 1984). He distinguishes "discursive consciousness," "practical consciousness," and "unconscious motives."<sup>2</sup> Deliberate or purposive thought and action are products of "discursive consciousness." The rational actor model takes deliberate action as the model for all action, and thereby gives discursive consciousness priority. Giddens recognizes a role for discursive consciousness, for example when an agent is engaged in carefully planned action. Under these circumstances, the contents of discursive consciousness are explanatory. However, because agency is prior to action, Giddens does not treat intentions, reasons, and motives as constitutive of action. The beliefs and attitudes cited as reasons for action are, more typically, elements of the way that agents reflect, *post hoc*, on their activity. They are means that agents use to justify their own actions and criticize the actions of others.

---

<sup>2</sup> Unconscious motives revolve around trust and anxiety, and Giddens develops them by appropriating ideas from the psychoanalytic theoretician Erik Erikson. While Giddens is right to suppose that an analysis of agency needs an affective dimension, and that the comfort of trust and the anxiety of shame are probably important, it would be much stronger to build on the basis of our best-supported cognitive theories of psycho-social development.

Since action is understood as an intervention in ongoing causal processes, the intentionality of action may be seen as the agent's capacity for monitoring the environment, his or her own responses, and the effects of those responses. Such "reflexive monitoring of conduct" is a part of the agent's practical consciousness. The agent is "conscious" of his or her own interventions into the causal flow of the environment in the sense that they are part of the agent's awareness, not in the sense that he or she is deliberating about them or focused on them. Giddens develops the idea of practical consciousness through a discussion of the role of memory in perception and action (Giddens 1984, 46ff). He highlights the way in which perception and action rely on established schemata. Perception is guided by activity, and it is through the engagement with particular objects that some aspects of the environment come to need special attention. The basic pattern of activity, the framework or schema for action, is learned by agents. Because of the pre-existing mental schema, agents can perceptually anticipate changes in their environment, categorize them, and respond by moving their bodies so as to have activity-relevant perceptual access.

Giddens distinguishes between memory—"the temporal constitution of consciousness"—and recall—"the means of recapitulating past experiences in such a way as to focus them upon the continuity of action" (Giddens 1984, 49). He then goes so far as to identify both discursive and practical consciousness with mechanisms of recall:

[D]iscursive and practical consciousness refer to *psychological mechanisms of recall*, as utilized in context of action. Discursive consciousness connotes those forms of recall which the actor is able to express verbally. Practical consciousness involves recall to which the agent has access in the *durée* of action without being able to express what he or she thereby 'knows.' (Giddens 1984, 49)

Giddens' use of recall places his conception of agency into a similar explanatory structure as Bourdieu's habitus. Agents learn to perceive and act within their social environment on the basis of already existing structures. Because their schemata of perception and action are the products of the current structure, their actions tend to reproduce it. Where Bourdieu had an unexplained explainer at the heart of his notion of agency, Giddens has filled in a mechanism. While this is a salutary and important advance on the problem, it leaves us with the same explanatory conundrum as Bourdieu. Reliance on mechanisms of recall explains the reproduction of practices while everything is going smoothly, but it is not clear how recall will help us explain how agents recreate structures as they break down. Giddens' treatment of agency as primary and his more detailed analysis are important steps, but they are not sufficient to resolve the conceptual difficulties with which we are confronted.

## 12.4 Agency as Ecological Attunement

None of the three conceptualizations of agency discussed above has adequate resources for explaining how improvising performers are able to create and maintain a musical structure through their joint action. Two lessons emerge from the previous section, one about structure and another about agency. Giddens and Bourdieu are unable to explain the recovery from a breakdown because they both conceptualize practices in terms of reproduction: practices are an enduring regularity of behavior which agents internalize and thereby reproduce. The explanatory problem set by the improvising musicians is not about the reproduction of the structure. The chord progression is repeated, but Monson does not ask: “In virtue of what is the sequence of chords in chorus 1 reproduced in chorus 2?” Indeed, the question is trivial in an ethnomusicological context. The microfoundational approach conceptualizes structure in more appropriate way. On that view, practices are not reproduced by agents, they are continually recreated. Existing patterns of behavior at a time are part of the environment for action. Agents recognize the patterns of action that arise in their social environment, just as they recognize patterns in the natural environment. Agents respond to that environment on the basis of their beliefs and utilities (and perhaps emotions, etc., as well), and the range of responses creates the new patterns. Existing practices thereby come to frame the possibilities for agents’ further interactions, but they are not reproduced by the agents. By not conceptualizing the problem of structure and agency as one of reproduction, it is much easier for the microfoundational approach to explain how practices change, and how agents might respond so as to maintain practices when they break down.

Practice theorists reject the microfoundational account because they reject its conceptualization of agency. The foregoing arguments further ground that rejection. Improvisation is not well represented as the product of choices. At the same time, none of the conceptions of agency have been adequate. The source of the problem, I suggest, is that all three treat agency as an individual phenomenon. The practice theorists’ framework of reproduction treats agents as social atoms: individuals internalize features of their social environment and then act on that basis. Microfoundational approaches notoriously do so as well, treating each agent’s decisions as a product of his or her internal representational states. Monson’s example of a breakdown and recovery of structure by an improvising group challenges these underlying commitments. The arguments above show that relying on internal states alone—whether they be intentions represented by signals or internalized structures—are insufficient to explain the group’s ability to recover from a structural breakdown. The central feature of Monson’s explanation is the mutual responsiveness of the players. The recovery was possible because the musicians acted in ways that changed the possibilities of action for the others, and each was responsive to the new possibilities. The agency illustrated in Monson’s account of improvisation is expressed by the musicians’ mutual responsiveness and coordination. If we take seriously Giddens’ suggestion that agency is

prior to action, then agency is partly constituted by the way each makes the others' actions possible. The second lesson, then, is that we need a way of conceptualizing agency as relational, not individual.

Combining these two lessons suggests that we should think of structures as being continuously sustained by mutual coordination and response, and of agency as the capacity for such interaction. The explanation of the recovery provided in Sect. 12.2 suggests how this capacity can be further analyzed. First, there are elements of conscious recognition and deliberate action. The piano player snapped his fingers, and then adopted a particular accompaniment style. Monson's interview with Byard shows that he was aware of the discrepancy between the piano and bass, and he was deliberately playing so as to rectify it. This aspect of her account is reminiscent of both Giddens' discursive consciousness and the rational action account of intentional action that invokes conscious beliefs and goals. The deliberate choices made by the musicians during this episode, however, are only a small part of Monson's story. To understand the adjustments, it was also crucial to understand the aspects of the environment to which the musicians were sensitive, and the possibilities for action they recognized for themselves and for each other. The description of what a chord progression is gives a taste of what a jazz musician is listening to as he or she plays. Their mutual responsiveness is possible because the musicians are listening to the same things—in this case the rhythms and the cycle of harmonies in the chord progression—and they practically understand how those elements of the environment create and limit possibilities for action. I want to suggest, then, that the mutual coordination and responsiveness manifested in musical improvisation is constituted by three features implicit in Monson's interpretation: what I will call meta-cognition, attunement, and affordance. Insofar as agency is the capacity for mutual coordination, agency can be analyzed to these three abilities.

The three elements implicit in Monson's explanation of jazz improvisation can be elaborated by relating them to known psycho-social cognitive capacities. *Attunement* to specific aspects of the environment is required to account for the improvisational maintenance of structure. This is, obviously, a matter of the psychology of perception, but it also depends on two specifically psycho-social cognitive mechanisms that have been prominent in recent psychological research. The first is the capacity to see some events as the intentional actions of an agent. This cognitive capacity has been demonstrated to arise very early in cognitive development and it arguably crucial for the development of a theory of mind (Meltzoff 1995; Bellagamba and Tomasello 1999). The second is the capacity for joint attention. We track the attention of others by noting their head and body position and their eye orientation to see that they are paying attention to the same things that we are attending to. Again, this plays a prominent role in a child's development, and it seems to be something that humans are particularly disposed to do (Moore and Dunham 1995; Carpenter et al. 1998). Practices are maintained by agents through their joint intentional actions. This means that agents must be able to track the same features of the environment (e.g. harmony and rhythm) and be aware that others are doing the same. Intentional actions are a crucial feature of any environment that includes human practices, so the capacities to recognize some actions as intentional



and to jointly attend to features of the environment are necessary conditions for the attunement of an agent who is participating in a social practice.

*Affordances* are possible actions available to an agent in an environment. These are not facts about the agent alone. Whether a window is an exit, for example, depends on both facts about the window (How big is it? How high is it?) and on facts about the agent (How big is he? How high can he jump?). An affordance is, therefore, a relation between the agent and the environment. In general, agency requires the capacity to recognize the affordances of a particular environment. In social action, agents must be sensitive to the possible range of action for both themselves and others. To maintain the harmonic and rhythmic structure of an improvised performance, each musician must attend to the relationship between the sounds he or she is currently producing and those produced by others. And each must anticipate the actions of others in their own current action. When the bass player begins the “obvious two-bar turnaround” in Monson’s example, the bass player has to recognize that one of the possibilities for action by the piano-player and drummer is to restart the chord progression on the downbeat that follows the turnaround. And if the structure is to be repaired, the piano player and drummer have to recognize this possibility as well.

Recognition of affordances requires a further cognitive skill that is specific to social environments. Humans have the capacity to represent the relationship between one’s own possibilities for action and the possibilities for others in a “third person” sort of way. Again, this competency is a part of normal human development, and children manifest it when they are three or four (Ashley and Tomasello 1998; Carpenter et al. 2005). This gives us the ability to easily trade roles in joint action. It also means that I can identify affordances of an environment that are not just my affordances: I can recognize the affordances for others. This permits us to recognize the possibilities for actions that require the contribution of more than one person. The agents’ understanding of their own and the others’ possibilities for action has to be a practical understanding. It is cognitively encoded in some way, of course, but that representation is closely linked to action and need not be accessible to conscious reflection or subject to deliberation. In Giddens’ terms, it would be part of practical consciousness. Such a capacity is clearly necessary for agents to participate in practices. Without the ability to recognize another person’s affordances, and see how those action possibilities fit with the agent’s own, coordinated action would be impossible.

The capacities for attunement and affordances capture the function of Bourdieu’s habitus, but with crucial differences: where the habitus replicates the structure, capacities for attunement and affordance are responsive and ecological. In Monson’s explanation, the harmonic and rhythmic structure of a jazz performance is maintained by the rhythm section through continuous monitoring of the actions of each other performer and its relation to the self’s current causal productions and the immediate possibilities for action. Since their instruments have different capacities and play different roles, they must recognize different affordances and have a practical understanding of how they interlock. More importantly, their actions not only influence each other, they partly constitute each other.

When the bass plays one note and the piano plays some others, they each provide part of the environment which changes the way the harmony sounds. Each contributes something that makes the other's action possible, and the harmonic and rhythmic structure arises as a consequence of this joint action. The capacities for attunement and recognition of affordance, then, are responsive and ecological in a way that the habitus is not. They are responsive in the sense that they are oriented toward the occurrent actions of the others in the immediate environment, rather than toward the pattern of practice. And they are ecological in the sense that the environment provides part of the content or substance of the possibilities for action.<sup>3</sup>

Attunements and affordances are at the level of Giddens' practical consciousness. They are psychological capacities that we do not normally bring to consciousness. They are dimensions of our ongoing causal interaction with the flow of events. Of course, we do act deliberately. Giddens' reflective monitoring provides a model of how the more conscious and deliberative side of our psychological capacities is related to the practical. The action of an individual should be understood as the joint product of three independent kinds of capacity: attunement to specific aspects of the environment, recognition of its affordances, and the *meta-cognitive* monitoring of ongoing action. The final element, meta-cognition, is our capacity to bring to mind prior plans, explicit beliefs about the environment, knowledge of explicit rules, and interpretations. When the piano player and drummer adopted a simpler accompaniment style, they were consciously and deliberately making a change for the sake of re-establishing the harmonic structure. Or at least, interview evidence suggests that the piano player did so. The drummer may have simply "followed" the piano, that is, have been sensitive to the change in the environment and its affordances. They did not, however, think through a plan or make a mini-composition in their minds with which to direct their action. Rather, the action was ongoing and maintained by their capacity to recognize and respond to aspects of their environment. The deliberate character was, as Giddens puts it, "causal intervention. . . in the ongoing process of events-in-the-world" (1979, 55).

## 12.5 Conclusion: Structure, Agency, and Improvisation

This essay argues for an ecological attunement conception of agency. It has three elements: the capacity to be attuned to the environment, to recognize the affordances for one's self and others, and to meta-cognitively reflect on one's interaction with the environment and thereby relate one's ongoing performance to

---

<sup>3</sup>In conversation, defenders of Bourdieu have responded that the habitus is supposed to be responsive to the environment in just this way. In other words, my interpretation of "habitus" is unduly restrictive. If so, then this argument can be interpreted as a way of making Bourdieu more precise and psychological plausible. Additionally, under this aspect, the argument points to a possible bridge between practice theory and cognitive anthropology.

prior plans, and propositional knowledge. This conception of agency is continuous with the work done by Bourdieu and Giddens, but it frees the conception of agency from the explanatory framework of reproduction. The characteristics of agents necessary for the maintenance of practice are capacities for response and coordination, not dispositions to absorb past regularities and reproduce them. Using the ecological conception of agency, we may think of structures as the microfoundational approaches do: as continuously recreated by agents' interaction. The ecological conception of agency adopts the microfoundational view of structure without the commitment to treating actions as paradigmatically deliberate. Even if some structures may be created and maintained by the interaction of strategic choices, clearly not all structures can be explained in this way. The ecological conception joins with Bourdieu and Giddens by enriching the conception of agency, and thereby strengthening its explanatory power.

The ecological conception of agency emphasizes the relational character of agency, and in this way it differs from all three conceptions discussed in Sect. 12.3. To be an agent is to stand in particular relationships to other agents and to the rest of the environment. The capacities of attunement and recognition of affordances are not representational states, but relationships between the organism and environment. To be attuned is to be attuned to something; to recognize an affordance is to have a practical understanding of what the object allows one to do. Thinking in terms of attunement and affordance lets the ecological conception of agency appeal to the stability of the environment when explaining the persistence of patterns of behavior. The specifically human capacities for the interactions that build social structure depend on human psycho-social cognitive abilities. In particular, these include the ability to distinguish intentional from unintentional action, the capacity for joint attention, and the ability to recognize the affordances for both self and other. To be an agent thus requires treating others as agents and responding to the joint possibilities for action. There could not be a world with just one agent. One person does not constitute a band, not because there is an insufficient number of bodies to institute practices, but because until she encounters another, she is not yet an agent.<sup>4</sup>

## References

- Ashley, J., & Tomasello, M. (1998). Cooperative problem-solving and teaching in preschoolers. *Social Development*, 7(2), 143–163.
- Banks, M. (2012). MacIntyre, Bourdieu and the practice of jazz. *Popular Music*, 31(1), 69–86.

---

<sup>4</sup>This essay is dedicated to the memory of Ivan Karp. Its development was facilitated by discussions not only with Ivan, but from critical commentary by Cory Kratz, Frank Lechner, Julie Zahle, and Fin Collin. Work on this essay was supported at the University of Hradec Králové in the Czech Republic by a Fulbright Scholar Grant and the Czech-US Fulbright Commission.

- Bellagamba, F., & Tomasello, M. (1999). Re-enacting intended acts: Comparing 12- and 18-month-olds. *Infant Behavior and Development*, 22(2), 277–282.
- Berliner, P. (1994). *Thinking in jazz: The infinite art of improvisation*. Chicago: University of Chicago Press.
- Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge: Cambridge University Press.
- Bourdieu, P. (1977). *Outline of a theory of practice*. Cambridge: Cambridge University Press.
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4), 1–174.
- Carpenter, M., Tomasello, M., & Striano, T. (2005). Role reversal imitation and language in typically developing infants and children with autism. *Infancy*, 8(3), 253–278.
- Elster, J. (2007). *Explaining social behavior: More nuts and bolts for the social sciences*. Cambridge: Cambridge University Press.
- Giddens, A. (1979). *Central problems in social theory: Action, structure, and contradiction in social analysis*. Cambridge: Cambridge University Press.
- Giddens, A. (1984). *Constitution of society: Outline of the theory of structuration*. Cambridge: Polity Press.
- Grice, P. (1989). *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Karp, I. (1986). Agency and social theory: A review of Anthony Giddens. *American Ethnologist*, 13(1), 131–137.
- King, A. (2000). Thinking with Bourdieu against Bourdieu: A ‘Practical’ critique of the habitus. *Sociological Theory*, 18(3), 417–433.
- Little, D. (1989). *Understanding peasant China: Case studies in the philosophy of social science*. New Haven: Yale University Press.
- Little, D. (1998). *Microfoundations, method, and causation: On the philosophy of the social sciences*. London: Transaction Publishers.
- Little, D. (2007). Levels of the social. In S. P. Turner & M. W. Risjord (Eds.), *Philosophy of anthropology and sociology*. Amsterdam: Elsevier.
- Lopes, P. (2000). Pierre Bourdieu’s fields of cultural production: A case study of modern jazz. In N. Brown & I. Szeman (Eds.), *Pierre Bourdieu: Fieldwork in culture*. Lanham: Rowman & Littlefield.
- Meltzoff, A. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31(5), 838–850.
- Monson, I. (1996). *Saying something: Jazz improvisation and interaction*. Chicago: University of Chicago Press.
- Moore, C., & Dunham, P. J. (Eds.). (1995). *Joint attention: Its origins and role in development*. Hillsdale: Lawrence Erlbaum.
- Schelling, T. C. (1969). Models of segregation. *American Economic Review*, 59(2), 488–493.
- Southern, E. (1997). *The music of black Americans: A history*. New York: Norton.
- Turner, S. (1994). *The social theory of practices*. Chicago: University of Chicago Press.

# Chapter 13

## Are Individuals Fickle-Minded?

Mathew D. McCubbins and Mark Turner

**Abstract** Game theory has been used to model large-scale social events—such as constitutional law, democratic stability, standard setting, gender roles, social movements, communication, markets, the selection of officials by means of elections, coalition formation, resource allocation, distribution of goods, and war—as the aggregate result of individual choices in interdependent decision-making. Game theory in this way assumes methodological individualism. The widespread conclusion that game theory predictions do not in general match observation has led to many attempts to repair game theory by creating behavioral game theory, which adds corrective terms to the game theoretic predictions in the hope of making predictions that better match observations. But for game theory to be useful in making predictions, we must be able to generalize from an individual’s behavior in one situation to that individual’s behavior in very closely similar situations. In other words, behavioral game theory needs individuals to be reasonably consistent in action if the theory is to have predictive power. We argue on the basis of experimental evidence that the assumption of such consistency is unwarranted. More realistic models of individual agents must be developed that acknowledge the variance in behavior for a given individual.

---

M.D. McCubbins  
Department of Political Science, Duke University, 210 Science Drive, Box 90362, Durham,  
NC 27708, USA

School of Law, Duke University, 210 Science Drive, Box 90362, Durham, NC 27708, USA  
e-mail: [mat.mccubbins@gmail.com](mailto:mat.mccubbins@gmail.com)

M. Turner (✉)  
Department of Cognitive Science, Case Western Reserve University, 607 Crawford Hall,  
10900 Euclid Avenue, Cleveland, OH 44106, USA  
e-mail: [mark.turner@case.edu](mailto:mark.turner@case.edu)

## 13.1 Introduction

Methodological individualism focuses on individual agents. It views large-scale social phenomena as the result of individual mental states that lead to actions. Specifically, in this book, methodological individualism is defined as the view that all explanations within the social sciences should be centered around individuals, their actions, beliefs, preferences, and the like. Accordingly, social phenomena such as the French revolution, an increase in the crime rate, residential segregation, the government's decision to lower the taxes, and the occurrence of unions are to be explained in terms of individuals, their actions, and the like.

Game theory analyzes outcomes as the aggregate result of choices that players make during interdependent decision-making. These choices are viewed as grounded in individual cognition by the players about the players involved, the actions they take, the information they possess, the strategies available to them, the outcomes they anticipate, and the equilibria that can be achieved across them.

Game theory makes a core set of assumptions about individual agents, namely, that they have consistent beliefs and preferences; that their actions result consistently from those preferences and beliefs; that these preferences, beliefs, and actions remain consistent across equal choice moments; and that the basic mental processes and inference procedures by which preferences and beliefs lead to actions remain the same under all conditions.

It is widely established experimentally that subjects do not in general follow the predictions of game theory. Accordingly, behavioral game theorists have stepped in with new assumptions about consistent *deviations* from classical rationality and assigned to subjects consistent dispositions to account for these deviations—dispositions having to do with risk preference, cognitive abilities, social norms, etc. All of these theories are fundamentally cognitive theories, making claims about how individual human minds work when choosing.

In this chapter, we assess the game-theoretic modeling of individual agents. We argue that these game-theoretic models assume a consistency in these agents that is, as yet, unwarranted. We argue that more realistic models of agents must be developed on the basis of systematic empirical research.

## 13.2 Generalizing from Experimental Data

The conception in game theory that consistent preferences drive decisions has had extraordinary influence lately. Literature in the social sciences is replete with publications reporting choices by subjects engaged in economic games in laboratory settings. Typically, these articles draw macroscopic inferences for real behavior from the behavior of these individuals in the laboratory. For example, here is the thesis of a highly-cited 1995 paper on general human behavior in transactions:

We designed an experiment to study trust and reciprocity in an investment setting. . . . Observed decisions suggest that reciprocity exists as a basic element of human behavior and that this is accounted for in the trust extended to an anonymous counterpart. . . . Is trust a primitive in economic models of behavior? What factors increase (or decrease) the likelihood of trust in economic transactions? We provide answers to these questions in a specific experimental setting, the investment game. By guaranteeing complete anonymity and by having subjects play investment game only once, we eliminate mechanisms which could sustain investment without trust; these mechanisms include reputations from repeat interactions, contractual precommitments, and potential punishment threats. We then show that positive investments still occur, suggesting that trust is an economic primitive. (Berg et al. 1995, pp. 122–123.)

The reasoning in this chapter—leading to the conclusion that trust is an economic primitive for individual agents—follows a common path: some subjects play a single game in a laboratory setting; their behavior is interpreted as the reliable outward sign of how they think and decide; accordingly their behavior in this one game in an experimental setting is used as a principle for modeling them as individual agents *tout court*; and this model of individual thinking and individual deciding is then generalized to human cognition and behavior in the world.

The ambition to use economic games run in laboratory settings as a microcosm of the world is understandable. Science prefers when possible to reduce vast complexity to simpler principles, to smaller pictures congenial to human thought. It would indeed be most useful if this reduction of social phenomena to a summation of actions by individual subjects in economic games proved to be scientifically legitimate.

But we should be cautious: the history of human ideas is replete with reductions that have turned out to be wrong, often to the surprise of generations of people who relied on them. In many cases, these misguided reductions are still powerfully with us. Vedic astrology continues to exert strong influence on decision-making among Hindus. The Tarot deck provides a remarkable microcosm for understanding the future and for planning accordingly, but it has not been shown to have any scientific value. Haruspication of entrails, augury, cartomancy, palmistry, pyromancy, and I Ching divination are similar reductions. These reduction strategies are often parodied—once they are discarded by a culture—for their vacuity.

In this paper, we present experimental evidence indicating that the drawing of macroscopic inferences about human behavior from the behavior of individuals in individual economic games in laboratory settings is not yet warranted.

### 13.3 What Is Going on in the Laboratory?

Laboratories in the social sciences are unlike laboratories in the physical sciences. Laboratories in the physical sciences are constructed under the view that the physical conditions in the laboratory can be designed so as to match exactly the relevant conditions of interest in the world. There is in this conception nothing, so to speak, about the bench scientist's laboratory bench *per se* that stops the scientist

from generalizing from what happens on the bench to the rest of reality, inclusive of those real situations that take place far from the scientist's bench. This happy conception of seamless generalizability allows the physical sciences, with the right care and nuance and adjustments, to claim that causal relationships detected in the laboratory for the most part generalize automatically to the world.

The case is utterly different for experiments on human beings, who belong to a social species evolved for behavior under certain conditions, and those conditions are not laboratory conditions. In principle, the burden is on the experimenter to show that the conditions in the laboratory do indeed match the relevant conditions of interest in the world in all the right ways. This can be a heavy burden, for several reasons.

We must assume as a beginning point that laboratory experiments involving economic games should fail to generalize to the world. The first reason we must make this defensive, defeasible assumption is that there are powerful and well-known "experimenter effects." For a physical or chemical system, or most biological systems, such as algae, no one imagines that the system is thinking, consciously or unconsciously, about the experimental situation and the experimenters. But human subjects are thinking, consciously and unconsciously, about both the experimental condition and the experimenters. Accordingly, absent compelling proof to the contrary, although one can assume that the data from such an experiment, if the experiment is impeccable on all other methodological scores, reveals something about how the subject behaves in that experimental condition, one cannot in principle assume that it reveals anything about how the individuals behave outside of that experimental condition.

The second reason for doubting that data from experiments involving economic games will generalize is that human beings and human cognition are evolved for messy environments. Conditions of the laboratory are sparse—which means that they are not the conditions for which human beings are evolved. The fact that the conditions of the laboratory are sparse requires us to be skeptical that generalization from experimental data to ecological human behavior is legitimate. Vision, for example, is evolved for conditions of white light, which are quite messy. Tests in the laboratory on human vision using simple, clean, monochromatic light do indeed show something important—namely how the human vision system operates under those experimental conditions—but it does not generalize to normal human vision. Color constancy, for example, an indispensable feature of human vision and inference, does not work under monochromatic light the way it does under normal conditions. An experiment under conditions that have not been demonstrated to match those in the real-world situations of interest is called "ecologically invalid," or just "invalid."

In sum, methodological rigor requires that we begin from the default assumption (albeit a defeasible default assumption) that laboratory experiments involving economic games do not generalize to human behavior. We are warranted in giving up that assumption only where high covariance has been reliably demonstrated between the behavior in the economic game and in the normal human setting.



It cannot in general be assumed that inferences established from the microcosm of the economic games in laboratory settings generalize to the human macrocosm.

### 13.4 Classical Economics and “Playing Nash”

Research using economic games in experimental settings begins with the baseline assumption—taken from classical economics, as in (Morgenstern and von Neumann 1947)—that subjects will optimize their payoffs within the strip of interdependent decision-making called the “game,” and do so by assuming that other agents will optimize their own payoffs. On these assumptions, interaction in interdependent decision-making must follow an equilibrium path. In this chapter, we will say, without summarizing the well-known details, that a game-player is “playing Nash” when she is following a rule for play (a “solution concept”) that will give her the optimal payoff that she can achieve through her own unilateral choices in a game where she is assuming both that all the other players are playing Nash and that all the players know each other’s equilibrium strategies. More generally, we will say that the classical paradigm proposes to explain human behavior through closed-form analytic models as a function of the Players involved, the Actions they can take, the Information they possess, the Strategies available, the Payoffs for actions, the Outcomes for the players, and the Equilibria that can be achieved across players—PAISPOE, for short.

Experiments with subjects playing economic games show that in general they do not play Nash. This is the oldest news on the planet, and we have nothing to add to that consensus, except that our batteries of experiments show the same thing. Details are available in (McCubbins and Turner 2012; McCubbins et al. 2012a, b, c).

### 13.5 Epicycles

Interpreters of data often guess why the players do not play Nash. To have scientific weight, these guesses would need to survive being tested as new hypotheses against out-of-sample data. Treating these guesses as knowledge would be “adding epicycles”—a slang term for “bad science.” The term refers to the supposed penchant of Ptolemaic astronomers to preserve the underlying theory by adding cycles-upon-cycles-upon-cycles as needed to erase the divergence between the theory and the known data.

The need to avoid epicycles in scientific investigation is well understood. Gigerenzer (2004, p. 602) offers what he calls “Feynman’s conjecture”:

To report a significant result and reject the null in favor of an alternative hypothesis is meaningless unless the alternative hypothesis has been stated before the data was obtained.

### 13.6 “Bounded Rationality” as an Epicycle

The first and still the best-known patch for PAISPOE models is “Bounded Rationality,” a term coined by Herbert Simon. According to Simon, rationality of the PAISPOE variety is limited because people lack information or have cognitive limits, including limits imposed by inability to think fast enough in the time available. Without a doubt, as every cognitive scientist knows, cognitive limits often make it impossible for people to do full PAISPOE calculation. Also without a doubt, lack of information can impede PAISPOE reasoning. Work by scholars such as Herbert Simon on “satisficing” and Gerd Gigerenzer on “heuristics” has contributed to our understanding of alternatives to PAISPOE reasoning.

Asserting that bounded rationality accounts for the mismatch between data and PAISPOE models is not in principle scientifically illegitimate. Quite the contrary. But the assertion is merely an epicycle if it is presented as an explanation for the mismatch, in the absence of a demonstration that a particular limit is indeed the cause of the mismatch.

When economists guess that subjects are failing to play Nash because bounded rationality impedes their ability to understand the structure of the game and its payoffs, the economists sometimes train the subjects on the game through “trial” rounds before they begin gathering the data that will be the basis of their conclusions. Training utterly stops any possibility of generalizing the behavior in the game to ecological behavior, for two reasons: (1) training creates an absolute difference between crucial conditions in the experiment and the ecological situations in which human beings have not been trained; (2) it is well-known in cognitive science that human beings can be trained to a frame that is contrary to their own patterns, and trained to it so well that it no longer seems alien; the benefits of such training are widely known in the martial arts, navigation, mathematics and scientific reasoning, diplomacy, and so on; and there is no reason to assume that behavior under training to a frame will generalize to normal human behavior—indeed, the mismatch was the very reason for training the human being.

Although it is indisputable that human rationality is bounded, adding “bounded rationality” to PAISPOE models as an epicycle has not provided us with any better models of human behavior than were provided by classical economics of the Morgenstern & von Neumann variety.

### 13.7 Framing

Since framing can influence decision, it is often proposed that deviation from Nash is accounted for by framing. The classic example of such a framing analysis is Kahneman and Tversky’s “Prospect Theory,” which proposes that differences in the framing of a choice can bias the choice one way or another despite the fact that the framing is immaterial to the consequences of the action with respect to the

payoff matrix (Kahneman and Tversky 1979, 2000). If we frame an action as a trade, then, since every trade is both a loss and a gain, it is possible for us to frame the action so as to emphasize loss or gain. Prospect theory proposes that there is a bias depending on this framing (Tversky and Kahneman 1992; Tversky and Fox 2000): it is assumed within expected utility theory that choosers are in general risk-averse, but, on the contrary, *ceteris paribus*, there is, according to prospect theory, a four-fold pattern of risk attitudes: risk-seeking for gains of low probability and for losses of high probability; risk-aversion for gains of high probability and for losses of low probability (Tversky and Fox 2000, p. 94). Accordingly, choosers will tend to make different choices depending on how the choice is framed—as loss or gain—despite the fact that the expected values of the alternative choices are identical. Kahneman & Tversky focus on framing effects in the decision-theoretic problem of choosing between alternative lotteries. Economists have since expanded this line of research into game-theoretic contexts, showing for example that framing affects the choice to contribute to a public good or impose externalities on others (see for example Andreoni 1995; Cookson 2000). McDermott et al. (2008) argue that “context-dependent” attitudes toward risk have a basis in evolutionary psychology. Post et al. (2008) show this same sensitivity to framing in the high-stakes choices of contestants on the game show “Deal or No Deal,” a decision environment decidedly far-removed from the foraging of our evolutionary ancestors.<sup>1</sup>

### 13.8 Character Type

It is often proposed that subjects have a certain character or psychological type that accounts for their deviation from Nash. For example, it is purported that people vary in the extent to which they are “self-regarding” versus “other-regarding.” Purportedly, people vary in their “risk preference.” Purportedly, people vary in their tendency to forego personal gain when doing so delivers a comparably much larger gain for other players. Purportedly, people vary in their preference for “fair” outcomes. And so on. It is also proposed that different players have different “level-k” signatures in particular settings. The idea behind “level-k” signatures is simple, and often used in films and novels. Consider, for example, this passage from *The Princess Bride*:

The Sicilian smiled and stared at the wine goblets. “Now a great fool,” he began, “would place the poison in his own goblet, because he would know that only another great fool would reach first for what he was given. I am clearly not a great fool, so I will clearly not reach for your wine.”

“That’s your final choice?”

---

<sup>1</sup> Unlike experimental studies of framing, Post et al. rely on observational data, in which the frame (previous earnings) is not controlled by an experimenter but generated endogenously by the subject.

“No. Because you knew I was not a great fool, so you would know that I would never fall for such a trick. You would count on it. So I will clearly not reach for mine either.”

“Keep going,” said the man in black.

“I intend to.” The Sicilian reflected a moment. (Goldman 1973, pp. 139–140.)

The Sicilian, to make his decision, is thinking that the man in black is thinking that the Sicilian is thinking that the man in black is thinking that the Sicilian is thinking that . . . In theories of Level-k reasoning, we begin with Level Zero. It is not clear to us from the literature what Level Zero is thought to be, but it is described as “unstrategic thinking,” so perhaps a Level-0 thinker (say Ann) simply shoots straight for the maximum payoff for herself in the payoff matrix, without any thought that the other player (say Paul) might have preferences of his own and so, strategically, interfere by making choices that move the Level-0 player, Ann, toward a different outcome, one that is better for Paul. “Unstrategic thinking” might mean that Ann chooses as if she is playing against random, non-intentional events—perhaps a role of the dice. She is then playing “against nature,” in the economic parlance, where, oddly, “nature” does not include intentional human cognition. Paul is a Level-1 thinker if he is playing so as to interact optimally with a Level-0 player. And so on up the line: a Level-2 thinker is imagining what a Level-1 thinker is thinking and responds optimally to a Level-1 thinker’s strategy. And a Level-3 thinker responds optimally against a Level-2 thinker, and so on.

The Sicilian, a self-assessed genius (“Never go in against a Sicilian when death is on the line!”) is many k-levels beyond everyday human subjects. He is even careful to prepare against potential adaptive behavior by the man in black (e.g., the man in black, an unequaled fighter, might try to kill the Sicilian, who is holding a large knife to Buttercup’s throat to prevent this adaptive behavior): the Sicilian distracts the man in black, managing to get him to look away from the goblets for a second, during which brief interval of time the Sicilian switches the goblets. After they have drunk and swallowed, the man in black announces that the Sicilian has guessed incorrectly. The Sicilian crows, “You only *think* I guessed wrong . . . That’s what’s so funny. I switched glasses when your back was turned.” Of course, the Sicilian dies in the next second from the iocane powder poison: the man in black, like human beings everywhere, has behaved adaptively rather than strategically in the game. As he explains to Buttercup, “They were both poisoned. I’ve spent the past 2 years building up immunity to iocane powder.”

Guesses can be hypotheses, but not explanations. Adding terms or factors to a theory that has failed tests, for the sake of making the theory fit the data, is methodologically acceptable if these changes are regarded as new, untested hypotheses.

### 13.9 A Battery of Experiments to Test Epicycles

To test whether behavior in economic games can be generalized to the world at all, we look for the most likely candidate, that is, behavior to which the generalization is most likely to apply. We take it that behavior in one economic game is most likely to apply to behavior

1. *by the same subject*
2. *under identical experimental conditions*
3. *in closely similar economic games*
4. *very near in time*

Accordingly, we must put the identical subject through a continuous battery of such games under the identical experimental conditions. Furthermore, in running this battery, we should

1. avoid training subjects, as discussed above, yet
2. test that they understand the payoff matrices and strategies of other players by quizzing them;
3. make framing as spare and general as practical; and
4. hold framing as consistent as possible across this battery of measures so as to avoid variation in behavior owing to variation in framing.

This is what we have done. Our battery included many economic games.

### 13.10 Can We Generalize from Behavior in Economic Games?

Subjects in our experiments are told that they are randomly paired at the beginning of every task with someone else in another room and that all their behavior is anonymous and private and that all the subjects have the same information. They receive no feedback on their play or indeed any information except in the few interactive games that they play at the end of the battery (e.g. Trust), where they must be told what the other player sent. They know that they are paid for every task according to how they perform, and that they will be paid at the end of the experiment in private by an assistant who will know only their number and the envelope of cash to be given to the person with that number. The following analysis considers data from 190 subjects for four economic games in our battery: Trust, Dictator, Donation, and Majority Public Goods. Of course, we do not use these misleading names in describing the experiment to subjects. These names are only for ease of reference.

Let us begin with the Trust game. Player 1 and Player 2 both begin with \$5. Player 1 can send any integer dollar amount to Player 2, including \$0. Whatever Player 1 transfers is tripled and given to Player 2. Then Player 2 can return any

integer dollar amount to Player 1, and the game ends. Notice that if Player 1 sends anything but \$0, then Player 2's pot of money becomes at least twice as large as Player 1's, maybe much larger. For example, if Player 1 sends \$1 to Player 2, then Player 1 is left at that point with \$4, but the \$1 sent is multiplied by 3, so Player 2 has \$8. "Nash" for Player 1 is to send \$0. Do our 190 subjects play Nash as Player 1 in Trust? Hardly. 105 of 190, or 55.3 %, send money as Player 1. This is just an example of the well-established fact that human subjects, informed that they are playing in an economic game with other human subjects, cannot be relied upon to play Nash. It is this fact, as we discussed above, that induces the proposal of epicycles.

105 subjects out of 190 received money as Player 2 in the Trust Game. Can we count on them to play Nash and return \$0? No. Given that they had every reason to view themselves as having been placed in an advantageous situation, can we count on them to be generous and not play Nash? No. 64 of these 105 subjects, or 61 %, play Nice and return money, and 41 of these 105 subjects, or 39 %, play Nash.

At this point, we can all feel the great temptation to "explain" these events by reducing the causality in the decision-making to personality: 61 % of these Subjects are "Nice" (or "other-regarding") and 39 % are not nice (or "self-regarding," or whatever)—we call them "Nash."

Our central point is that such a conclusion depends upon assuming that human beings are consistent in their preferences and methods of making choices and that if they make a different choice it is because they face different conditions, yet it is just this assumption that most needs to be tested empirically.

Our battery of experiments was designed so as to let us test the hypothesis that "personality type" generalizes, in other words, that people are consistent. Does a subject's supposed "Nice" versus "Nash" type generalize to even the identical subject's behavior in identical settings and identical conditions during the same span of a few hours in a similar economic game? Our battery was designed so that subjects played both Player 2 in the Trust and Player 1 in Dictator *under the identical payoff conditions and, going forward, the identical game structure*. In the Dictator game, there are two players: the Dictator (Player 1) and the Receiver (Player 2). The Dictator has an endowment and chooses what part of it, if any, to send to the Receiver. The Receiver receives the amount sent and the Dictator keeps that part of the endowment the Dictator chose not to send, and that is the end of the game. The Receiver's role is entirely passive. We arranged our Dictator game so that the Dictator has the same endowment he or she has in the role of Player 2 in Trust, and that the Receiver has the same endowment he or she has in the role of Player 1 in Trust. These endowments are common knowledge. Accordingly, our Dictator game is identical to the second half of our Trust game. In effect, each individual subject plays the second half of the Trust game twice. Formally, there was no mathematical or economic difference in any individual subject's conditions as Player 2 in Trust and Player 1 in Dictator.

Specifically, for any specific subject  $S^*$ ,  $S^*$  was in the role of Player 2 in Trust at one point in the battery and Player 1 in Dictator at another point in the battery. We introduce the label  $t(S^*)$  for the other subject with whom  $S^*$  was randomly paired in

Trust. We introduce the label  $d(S^*)$  for the other subject with whom  $S^*$  was randomly paired in Dictator. In Trust, where both  $t(S^*)$  and  $S^*$  begin with \$5,  $t(S^*)$  sends an amount (perhaps \$0), which is tripled and added to the \$5 endowment which  $S^*$  had at the beginning of the game. At that point in the Trust game,  $S^*$  has  $a$  dollars and  $t(S^*)$  has  $b$  dollars. Later in the battery of experiments,  $S^*$  plays Player 1 in Dictator, and we arranged the experiment so that in Dictator, the endowment for that particular subject  $S^*$  is exactly  $a$  and the endowment for specific subject  $d(S^*)$  is exactly  $b$ . That is, the endowments that a given subject  $S^*$  faces in the two games are identical at these two points in the two games. Here is a table:

Money that players have after the P1 send in Trust:

$$S^* \text{ has } a \mid t(S^*) \text{ has } b$$

Money that players have as endowments in Dictator:

$$S^* \text{ has } a \mid d(S^*) \text{ has } b$$

At this point in each of the two games,  $S^*$  (for the 105 subjects who received money as Player 2 in Trust) has at least twice as much money as the other person. Accordingly, in Dictator, for these 105 subjects,  $a$  is always at least twice as large as  $b$ , and sometimes much bigger.

In both games, at this point, there is only one choice left to make, and that choice is identical in both games:  $S^*$  must choose how much, if any, of  $a$  to send to the subject with whom  $S^*$  is randomly paired in that game. So at this point in the two games, going forward, the two games have the identical structure and payoffs.

Do the purported Nice v. Nash personality types we might think we see when subjects play Trust Player 2 generalize even to the identical economic situation with the identical choice to make, now in Dictator?

No. 41 of the 64 Nice types as Trust Player 2 are Nice in Dictator, but 23 are Nash. So the generalization works for only 64 % of the Nice subjects. 37 of the 41 subjects who are Nash during Trust Player 2 are Nash as Dictator Player 1; the generalization on “Nash” personality type holds (at this point) for 90 % of subjects, making the “Nash” generalization look (at first blush) better than the “Nice” generalization, but still not a reliable generalization, since 10 % of the Nash-types in Trust Player 2 are Nice as Player 1 in Dictator.

Next, we compare the behavior of each of these subjects in Trust to the behavior by the same subject as Player 1 in the Donation game. In Donation, both players begin with \$5. Player 1 can pass any amount of the \$5 to Player 2. The amount is multiplied by 4 before it is given to Player 2. Then the game ends. In this case, any amount of Niceness by Player 1 results in a fourfold level of Niceness received by Player 2, as measured in money. For example, if Player 1 passes \$1, Player 1 is left with \$4, but Player 2 now has \$9. If Player 1 passes \$5, Player 1 is left with \$0, but Player 2 now has \$25. Do the Nice versus Nash personality types we think we might

see in Trust Player 2 and Dictator generalize subject by subject to the Donation Game?

No. Of the 41 most completely confirmed Nice types in Trust and Dictator, 27 % suddenly are Nash in Donation. That is, 41 of the 190 subjects receive money as Player 2 in Trust, send money as Player 2 in Trust, and send money as Player 1 in Dictator. But 27 % of those 41 are Nash in Donation! Now consider the 23 Ss who receive money as Player 2 in Trust, play Nice as Trust Player 2, but play Nash as Dictator Player 1. What do they do in Donation? Half (12) of those 23 play Nice and half (11) play Nash. Consider the 37 subjects of 105 with the clearest Nash character type: they receive money as Trust Player 2 yet return nothing, and also send nothing as Dictator Player 1. Can we at least count on this confirmed 35 % of the pool of 105 subjects who received money as Trust Player 2 to be rock solid Nash? No. In Donation, 30 % of them play Nice. And so on.

Next, we compare what these specific subjects did when they played the Majority Public Goods game. In this game, each subject is assigned randomly to a group of 10 subjects (about whom they know nothing and with whom they cannot communicate) and given \$5. The subjects can each keep the \$5 or move the \$5 to a group pot. If at least 6 of the 10 do so, then the pot is tripled and each subject in the group receives a 10 % share of the pot. If fewer than 6 of the 10 do so, then nothing from the pot is given back to the subjects. This game is not a perfect distinguisher between Nice and Nash, because there is one place where they overlap. A subject who plays Nash will not contribute if the subject thinks that 0, 1, 2, 3, 4, 6, 7, 8, or 9 of the other players in the group will contribute. But if the subject believes that exactly 5 of the others will contribute to the pot, then the subject believes that joining the group of givers would raise its membership to exactly 6, in which case the subject receives \$9 by contributing but \$5 by not contributing. Can we rely on the purported Nice types to play Nice in the Majority Public Goods game?

No. For example, of the 30 subjects out of 190 who receive money as Player 2 in Trust, return money as Player 2 in Trust, send money as Player 1 in Dictator, and send money as Player 1 in Donation, 13, or 43 %, do not contribute in the Majority Public Goods Game. Similarly, of the 26 subjects who receive money as Player 2 in Trust, return 0 as Player 2 in Trust, send 0 as Player 1 in Dictator, and send 0 as Player 1 in Donation, 5, or 19 %, contribute in the Majority Public Goods Game. Similarly, for other sub-sub-sub-subcategories of the subjects, we find that a putative “personality” signature is unreliable in predicting behavior in the Majority Public Goods Game.

In summary, of the 105 subjects who received money as Player 2 in Trust, only 17, or 16 %, keep a consistently “Trusting” or “Cooperative” or “Generous” signature, and only 21, or 20 %, keep a consistently “Ungenerous” signature.

But what about those 105 subjects who *send* money as Player 1 in Trust? Surely they were Nice. In the Trust game, both players do much better if they trust each other: if Player 1 sends the full \$5 available, then Player 2 has \$20 and can send \$10 back to Player 1. Both players then have doubled their initial endowment. Husbands and wives in community property states, or any two people under a trusted contract according to which they split the benefit, should, under Nash, immediately send



everything as Player 1 in Trust, because the contract means that you do not have to rely on the generosity of the other person: you own a 50 % share of all assets, and so does the other player.

Let us compare behavior in the Trust Game with behavior in the Prisoner's Dilemma game (PD). Prisoner's Dilemma was another part of our battery. It is always set up so that a strategy of cooperating is strictly dominated by a strategy of defecting: Whichever choice the other player makes, the subject is always better off in choosing to defect. A Nash player, of course, must defect. Yet, if both players cooperate, they are better off than if both players defect.

The data are complicated at this point by the fact that we tested four different kinds of framing of the identical choice and payoff structure. The subjects did not all receive the same framing. Methodologically, we might prefer then not to lump them together, but this raises an interesting point: we often see in the literature data lumped together from different experiments with different protocols, run at different times by different experimenters, moreover using a between-subjects design. In our case, we have a within-subjects design, in the sense that the same subjects played both Trust and Prisoner's Dilemma. Under all four Prisoner's Dilemma framings, the payoffs were identical and everything else was held constant, except for the four framings. In all four versions, each subject had the choice to cooperate or to defect. Lumping these four groups together is not methodologically clean, but it is fairly conservative relative to the practices we often find in the literature, and our overarching point here is that one should doubt generalizations assumed in the literature. Can generalizations hold up over this lumping? Here are the results.

We start with the 105 Ss who play Nice as Player 1 in Trust. Did they play Nice in Prisoner's Dilemma? No. 34 of 105, or 32 %, play Nice. But 68 % play Nash.

Now let us look at the other 85 subjects. 85 subjects play Nash in sending \$0 as Player 1 in Trust. Do they play Nash in Prisoner's Dilemma? Not so much. 65 of 85, or about 76 %, play Nash. But 20, or about 24 % play Nice.

### 13.11 Characterizations Do Not Hold Up

Perhaps there are other ways to use economic games as laboratories in which behavior can be generalized to the world. Perhaps there are other signatures, other reductions, in the form of characterizations. Perhaps there are other ways in which economic games can serve as a microcosm from which we can learn about the macrocosm of human behavior. But the ways we review here—all of which are forms of characterizing actors as having stable preferences and stable ways of making choices—do not withstand our tests of their validity, and conclusions from them should be held in abeyance until science develops more realistic models of individual agents on the basis of systematic empirical research.

### 13.12 Conclusion

Common sense tells us that people have beliefs and desires, or beliefs and preferences; that they are aware of them; and that they act according to them. But cognitive science has undermined commonsense notions of the mind. What we take for granted about human thought has proved to be unimaginably more complex than anyone had expected; to be profoundly misrepresented by our supposedly bedrock, commonsense, intuitive notions; and to be conducted almost entirely in the backstage of cognition, invisible to consciousness. The cartoons of consciousness are highly useful, and there is no evolutionary advantage in building consciousness so that it can see through them. Human beings are awesomely effective, but for the most part clueless about how they work.

The basic assumptions about the human mind made in PAISPOE models may seem unassailable, sheer common sense, but that cuts no scientific ice. Their status as common sense is no reason to accept them. Rather, they are hypotheses, and, as such, must be tested to have weight.

More than 30 years ago, Lee Ross (1977) coined the term “fundamental attribution error” for the excessive tendency of everyday “intuitive psychologists”—that is, everybody—to explain other people’s behavior by attributing dispositions to them. Jocularly but also aggressively, and certainly influentially, he proposed that this error was the main basis for the field of social psychology.

There are many assertions made in economics that depend upon the folk-psychology assumptions we see in PAISPOE models, such as that what players are doing under a set of beliefs must be in equilibrium. Where the data diverge from the PAISPOE models, it is tempting to deploy the fundamental attribution error to explain that divergence between model and data, by adding an epicycle that consists of characterizing the actors by attributing stable dispositions. But perhaps the PAISPOE assumptions, and the assertions that depend upon them, are wrong in the first place. In cognitive science, commonsense notions of how vision works, how language works, how memory works, how categorization works, how inference works, and so on have all fallen by the wayside.

We propose that the future of economics lies not in drawing further conclusions from PAISPOE assumptions but rather in testing them scientifically in order to recast the foundations of the field.

If we accept the game-theoretic assumption that people have consistent preferences and consistent methods of deciding and that different choices are the result of different conditions, what are we to make of the data from our battery of experiments? One hypothesis would be that we somehow fielded a group of fickle-minded people, alien individuals, who gave us data that we must throw out as bizarre, or that the experiments are corrupt, or that the design involved confounds, and so on—and all of these possibilities should be considered.

But there is another logical possibility that we must also consider: perhaps the assumption that individual agents have consistent preferences and consistent methods of decision-making that run across different situations and contexts is

wrong. After all, it is not clear on evolutionary or cognitive grounds that individual agents should be expected to work in this way. There is room for doubt. These assumptions of game theory may be reductions that we must discard. Before we launch centuries of research on the assumption that the solar system is geocentric, we should collaborate to take that assumption and test it to destruction. Before we launch centuries of research on the assumption that individual agents are to be modeled as consistent modulo circumstances, we should collaborate to take that assumption and test it to destruction. Our point is not at all that, with a little data, we prove that these assumptions are clearly wrong, but rather that we can now see that they are assumptions. We cannot base science on untested assumptions. If we are to build a house, we must build it on rock rather than sand, and if these principles are what we mean to build our house upon, we must first prove that they are rock and not sand. We have not done that.

This paper questions one prominent example of a theory committed to the thesis of methodological individuals, that is, game theory. We have argued that science needs a more adequate model of the individual actors than the one espoused by game theory. We need more realistic conceptions of agents—conceptions which must be developed on the basis of systematic empirical research. We do not suggest that a more adequate model of agents would automatically serve as support for the strong view of methodological individualism. No matter where the debates between methodological individualists and holists may land, there will be ample room in any workable social science for accounts that refer to actors.

## References

- Andreoni, J. (1995). Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments. *Quarterly Journal of Economics*, 110(1), 1–21.
- Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, 10, 122–142.
- Cookson, R. (2000). Framing effects in public goods experiments. *Experimental Economics*, 3, 55–79.
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33, 587–606.
- Goldman, W. (1973). *The Princess Bride, S. Morgenstern's classic tale of true love and high adventure, the 'good parts' version, abridged by William Goldman*. New York: Ballantine Books.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. Cambridge: Cambridge University Press.
- McCubbins, M. D., & Turner, M. (2012). Chapter 14: Going cognitive: tools for rebuilding the social sciences. In R. Sun (Ed.), *Grounding social sciences in cognitive sciences* (pp. 387–414). Cambridge, MA: MIT Press.
- McCubbins, M. D., Turner, M., & Weller, N. (2012a). The theory of minds within the theory of games. In *Proceedings of the 2012 international conference on artificial intelligence*. Las Vegas: Nevada.

- McCubbins, M. D., Turner, M., & Weller, N. (2012b). The mythology of game theory. In *Proceedings of the international conference on social computing, behavioral-cultural modeling, & prediction* (Springer lecture notes in computer science). College Park: University of Maryland.
- McCubbins, M. D., Turner, M. & Weller, N. (2012c). The challenge of flexible intelligence for models of human behavior. In *Technical report of the Association for Advancement of Artificial Intelligence spring symposium on game theory for security, sustainability and health*. Stanford, California: Stanford University.
- McDermott, R., Fowler, J. H., & Smirnov, O. (2008). On the evolutionary origin of prospect theory preferences. *Journal of Politics*, 70(2), 335–350.
- Morgenstern, O., & von Neumann, J. (1947). *The theory of games and economic behavior*. Princeton: Princeton University Press.
- Post, T., Van den Assem, M. J., Baltussen, G., & Thaler, R. H. (2008). Deal or no deal? Decision making under risk in a large-payoff game show. *American Economic Review*, 98(1), 38–71.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in Experimental Social Psychology*, 10, 173–220.
- Tversky, A., & Fox, C. R. (2000). Weighing risk and uncertainty. In D. H. Kahneman & A. Tversky (Eds.), *Choices, values and frames* (pp. 93–117). Cambridge: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5, 297–323.

# Index

## A

Abbott, A., 62–64, 67  
Adorno, T., 215  
Ainsle, G., 150  
Alexander, J.K., 148  
Allison, G., 158  
Ansell, C., 148  
Archer, M., 50, 62, 178  
Aristotle, 84

## B

Baltussen, G., 243  
Barker, E., 90  
Becker, G., 149  
Beckert, J., 69, 70  
Bhaskar, R., 46, 47, 62, 178  
Boudon, Raymond, 1, 10  
Bourdieu, P., 62, 220  
Bradley, F., 214  
Broad, Charlie, D., 46  
Buckle, H.T., 79

## C

Callon, M., 205  
Coleman, J., 58, 69, 117  
Collier, A., 62–64, 67  
Collin, F., 194, 197–216  
Collins, H., 206  
Comte, A., 79  
Correia, F., 23  
Craver, C., 123–125  
Currie, G., 21, 119

## D

Demeulenaere, P., 163  
Dennett, D., 212, 213  
Descartes, R., 84  
Dewey, J., 63, 64  
Doris, J., 148  
Durkheim, É., 79, 80

## E

Elder-Vass, D., 39–52, 58, 125, 178, 179, 181,  
182, 187  
Elster, J., 67, 144, 161, 221, 226  
Epstein, B., 17–37, 107, 121, 212, 215  
Erikson, E., 229

## F

Fodor, J., 213  
Fowler, J., 243  
Freeman, J., 145, 146, 151

## G

Garfinkel, A., 156  
Garfinkel, H., 61  
Ghaemi, N., 170  
Giddens, A., 62, 220, 229–231  
Gierke, O., 90  
Gigerenzer, G., 241, 242  
Gode, D., 149  
Goffman, E., 61  
Granovetter, M., 62, 64, 68–69  
Grice, P., 226  
Gross, N., 62, 64–68

**H**

Hacking, I., 78  
 Hannan, M., 145, 146, 151  
 Harré, R., 46  
 Hart, H.L.A., 29  
 Hayek, Friedrich, 1, 7  
 Hedström, P., 67, 113  
 Hobbes, T., 84, 89  
 Hodgson, G., 21, 23–25  
 Horkheimer, M., 215  
 Hume, D., 29

**J**

Joas, H., 62, 64, 65, 69–70

**K**

Kahneman, D., 242, 243  
 Kelly, D., 149  
 Kilpinen, E., 64  
 Kim, J., 19, 27, 28, 44, 45, 119–121, 125, 180, 186, 191  
 Kincaid, H., 18, 112, 119, 139–151, 155, 156, 168, 184  
 Korsgaard, C., 61, 64  
 Kripke, S., 87

**L**

Latour, B., 198, 200–204, 206, 208, 209, 211, 212, 216  
 Lawson, T., 178  
 Lewis, D., 108, 110  
 List, C., 90, 93, 94, 103, 104, 106–112  
 Little, D., 55–73, 122, 166–168, 184, 213, 220, 226  
 Lloyd, C., 165  
 Lukes, S., 21, 79, 184

**M**

Mallon, R., 149  
 Mandelbaum, M., 184  
 Marx, K., 201  
 McCubbins, M.D., 237–251  
 McDermott, R., 243  
 McTaggart, J., 214  
 Mead, G.H., 63, 69  
 Mitchell, S., 168–171  
 Monson, I., 223–225  
 Moss, L., 142

**N**

Nagel, E., 232, 233

**O**

Oppenheim, P., 121, 186

**P**

Padgett, J., 148  
 Parsons, T., 199  
 Pettit, P., 26, 28, 77–95, 98–107, 112–114, 119, 125, 128, 155, 166  
 Popper, Karl, 1, 4, 7  
 Post, T., 243  
 Putnam, H., 26, 121, 186

**R**

Risjord, M., 156, 212, 219–235  
 Rorty, R., 65, 66  
 Ross, L., 250  
 Rousseau, J.-J., 84, 85  
 Russell, B., 214

**S**

Sartorio, C., 107, 113  
 Sawyer, K., 163  
 Sayer, A., 178  
 Schelling, T., 221  
 Schweikard, D., 113, 114  
 Searle, J., 28, 31, 41, 42, 155  
 Shapiro, L., 110  
 Shockley, K., 98, 99, 101, 103–105  
 Simon, H., 242  
 Smirnov, O., 243  
 Smith, A., 79  
 Sober, E., 110, 144  
 Spiekermann, K., 112  
 Steinmetz, G., 71  
 Stephan, A., 46  
 Sun, R., 146  
 Sunder, S., 149  
 Szigeti, A., 97–114, 194

**T**

Thaler, R., 243  
 Thatcher, M., 146  
 Tolstoy, L., 79  
 Turner, J., 69

Turner, M., 213, 237–251  
Tversky, A., 243

**U**

Udehn, L., 154, 183

**V**

Van Bouwel, J., 153–173, 213  
Van den Assem, M., 243

**W**

Watkins, J., 142  
Weber, Max, 1  
Whitehead, A., 214

Whorf, B., 59  
Wimsatt, W., 28, 127  
Wittgenstein, L., 87  
Woodward, J., 103, 108–110, 123, 124,  
126, 131  
Woolgar, S., 201–204

**Y**

Yearley, S., 206  
Ylikoski, P., 100, 117–133, 154, 166, 167, 192,  
194, 214, 215

**Z**

Zahle, J., 24, 40, 107, 112, 118, 119, 125, 143,  
154, 162, 177–194, 213, 215