

DE GRUYTER

*Sonja Rinofner-Kreidl,  
Harald A. Wiltsche (Eds.)*

# ANALYTIC AND CONTINENTAL PHILOSOPHY

PROCEEDINGS OF THE 37TH INTERNATIONAL  
WITTGENSTEIN SYMPOSIUM

PUBLICATIONS OF THE AUSTRIAN  
LUDWIG WITTGENSTEIN SOCIETY – N.S.

## **Analytic and Continental Philosophy**

**Publications of the Austrian  
Ludwig Wittgenstein Society  
New Series (N.S.)**

—

**Volume 23**

# **Analytic and Continental Philosophy**

---

Methods and Perspectives. Proceedings  
of the 37th International Wittgenstein Symposium

Edited by  
Sonja Rinofner-Kreidl, Harald Wiltsche

**DE GRUYTER**

ISBN 978-3-11-044834-4

e-ISBN (PDF) 978-3-11-045065-1

e-ISBN (EPUB) 978-3-11-044887-0

ISSN 2191-8449

**Library of Congress Cataloging-in-Publication Data**

A CIP catalog record for this book has been applied for at the Library of Congress.

**Bibliographic information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.dnb.de>.

© 2016 Walter de Gruyter GmbH, Berlin/Boston

Printing: CPI books GmbH, Leck

♻️ Printed on acid-free paper

Printed in Germany

[www.degruyter.com](http://www.degruyter.com)

# Table of Contents

Sonja Rinofner-Kreidl & Harald A. Wiltsche

**Introduction — 1**

## Wittgenstein

Peter Hacker

**Can You Have My Pain? — 11**

Martin Kusch

**Wittgenstein's *On Certainty* and Relativism — 29**

Christian Helmut Wenzel

**Wittgenstein and Free Will — 47**

Joachim Schulte

**Wittgenstein's Last Writings — 63**

## Metaphilosophy and Methodology

Dan Zahavi

**Analytic and Continental Philosophy: From Duality Through Plurality to (Some Kind of) Unity — 79**

Charles Siewert

**For Analytic Phenomenology — 95**

Steve Fuller

**Towards a New Foundationalist Turn in Philosophy: Transcending the Analytic-Continental Divide — 111**

Guillaume Fréchette

**Two Phenomenological Accounts of Intuition — 129**

Julia Jansen

**Imagination and 4E Cognition: An Analytic-Continental Exchange — 143**

Thorsten Streubel

**Intuition und Argumentation – zum Verhältnis von intuitiver und diskursiver Vernunft — 157**

Uwe Meixner

**Truth against Reason, and Reason against Truth — 173**

## Philosophy of Mind

Jan Slaby

***Don't beep me, bro!* – A Worry About Introspection — 187**

James N. McGuirk

**Metaphysical and Phenomenological Perspectives on Habituality and the Naturalization of the Mind — 203**

Thomas Fuchs

**Embodied Knowledge – Embodied Memory — 215**

Michel Bitbol

**Panpsychism in the First Person — 231**

Karl Mertens

**What Is It Like to Be an Angel?**

The Importance of Corporeally-Situated Experience for Our Concept of Action — 247

## Social Philosophy and Collective Intentionality

Thomas Szanto

**Do Group Persons have Emotions – or Should They? — 261**

Alessandro Salice

**Collective Intentionality and the Collective Person in Max Scheler — 277**

Hans Bernhard Schmid

**Being Well Together – Aristotle on Joint Activity and Common Sense — 289**

Michela Summa

**Pretence and the Inner. Reflections on Expressiveness and the Experience of Self and Other — 309**

Hanne Jacobs

**Socialization, Reflection, and Personhood — 323**

Robert Hugo Ziegler

**'Vaulting Ambition' – Machiavelli's Empty and Impure Concepts — 337**

## Ethics and Value Theory

Marcia Baron

**The Distinction between Objective and Subjective Standards in the Criminal Law — 351**

Inga Römer

**Gibt es einen kantianischen Intuitionismus in der Ethik? — 369**

Roberta De Monticelli

**Sensibility and Values Toward a Phenomenological Theory of the Emotional Life — 381**

Kathi Beier

**Stolz und Vorurteil. Über einige Schwierigkeiten der ethischen Selbstbewertung — 401**

**Index of Names — 413**

**Index of Subjects — 419**





# Sonja Rinofner-Kreidl & Harald A. Wiltsche

## Introduction

Much ink has been spilled over attempts to specify the nature and significance of the infamous distinction between continental philosophy (henceforth: *CP*) and analytic philosophy (henceforth: *AP*). In the eyes of some, the *AP/CP* distinction does not merely reflect gradually different orientations within one discipline. Rather, *CP* and *AP* are sometimes said to be more like incommensurable paradigms, each constituting fundamentally distinct modes of thought that have little, if anything, in common. According to Michael Dummett, for instance, “[i]t’s no use now shouting across the gulf [because] we have reached a point at which it’s as if we’re working in different subjects” (Dummett 1993, p. 193). More recently, Anthony Quinton has expressed similar sentiments, writing that there is “no perceptible convergence between the two philosophical worlds” (Quinton 2005, p. 172). And A.J. Ayer’s autobiographical remarks about his interactions with Maurice Merleau-Ponty seem to corroborate both Dummett’s and Quinton’s views:

We did [...] attempt it [to find some common ground for philosophical discussion] on several occasions, but we never got very far before we began to wrangle over some point of principle, on which neither of us would yield. Since these arguments tended to become more acrimonious, we tacitly agreed to drop them and meet on a purely social level, which still left us quite enough to talk about. (Ayer 1977, p. 285)

If we take all this at face value, then contemporary philosophy indeed appears to be “A House Divided” (cf. Prado 2003). However, as our cautious tone already indicates, we are not entirely convinced that this is really the case.<sup>1</sup> Perhaps

---

**1** It is interesting to note, however, that those who believe in the existence of two incommensurable philosophical paradigms hardly ever offer arguments as to why this should necessarily be a bad thing. Looking at the history of the sciences, for instance, we find several examples of conflicting paradigms that battled for ascendancy and definatory power at the same time. In many cases, the differences extended well beyond the level of theory-construction and involved fundamental questions concerning methodology, standards of rationality or divergent views about what constitutes scientific evidence, thus making cross-paradigmatic communication a real challenge. Would anyone be surprised if, say, Galileo had complained that he and his Aristotelian opponents “have reached a point at which it’s as if we’re working in different subjects”? Of course, the point is not to draw conclusions about philosophy on the basis of the historical record of the sciences. The point rather is that competing frameworks can instill a fruitful dynamics in pondering philosophical issues and can push ahead philosophical debates even on condition that the opposing parties do not directly interfere with one another. Though framework transgressing recognition of peers is no trivial problem that deserved separate dis-

the picture of two neatly separated camps is, at least in part, an illusion or, as one nowadays prefers to say, a collective narrative. Of course, it makes sense to ask for the functions such narratives meet. In our view, there is overwhelming evidence that these are primarily institutional and practical functions in a wider sense. To draw a clear-cut distinction between AP and CP may help, for example, in understanding how academic networking, research dissemination strategies and job markets function. But there are good reasons for being skeptical about whether the distinction bears much philosophical impact. It is remarkable, for instance, that although the distinction is still widely used, no one seems to know how to draw it in a sufficiently clear and non-arbitrary way. As a former organizer of the International Wittgenstein Symposium has rightly observed, one feels reminded of Saint Augustine's puzzlement about the concept of time (Reicher 2005, p. 12): As long as we are not asked what it is, we seem to understand the AP/CP-distinction perfectly well; but if we are pressed to explain it in detail, we find it hard to give a satisfactory answer.

In our experience, philosophy students are able to determine the *extensions* of the terms "AP" and "CP" at a relatively early stage of their studies. AP is, they seem to learn, the kind of philosophy we find in the writings of Frege, Russell, Carnap, Quine, Davidson, Kripke or Anscombe. And, in the vast majority of cases, it is also the sort of philosophy that is encountered when browsing through the latest issues of *Mind*, *Philosophical Studies* or *Nous*. CP, on the other hand, is what students are told to find in the writings of Heidegger, Sartre, Jaspers, Gadamer, Derrida, Foucault, Deleuze or Butler. There are, to be sure, tricky cases such as Husserl, Rorty, Feyerabend or Wittgenstein. But, by and large, it is, it seems, not too much of a challenge even for undergraduates to put a given philosopher in one of the two available boxes.

But what are the grounds on which these (or slightly different) classificatory schemes work? We all know that there exist several well-known stereotypes about what it takes to belong to either of the two camps: Depending on one's own preferences, AP is characterized by its rigour, its clearness, its level of sophistication, or by its dullness, its artificiality and its remoteness from real-life problems. CP, on the other hand, is characterized by its concreteness, its relevance for fields outside of philosophy, its rootedness in the life-world, or by its obscureness, its lack of precision and its hostility to science. But it is clear that none of these characterizations is sufficiently neutral or precise to serve as a basis for a classificatory practice that is worthy of a discipline like philosophy.

---

cussion, it is indeed sometimes enough to notice an opponent's radically different views to feel pressed to work hard on one's own arguments.

Perhaps it is more promising to distinguish between AP and CP on the basis of the topics they typically address. According to Simon Critchley, for instance, CP is “marked by a strong consciousness of history” (2001, p. 57). But, although there is certainly some truth in this, consciousness of history is hardly exclusive to thinkers associated with CP. As early as in 1962 Norwood Russell Hanson claimed that, while “history of science without philosophy of science is blind, [...] philosophy of science without history of science is empty” (1962, p. 580). In the meantime, after the pioneering work of Imre Lakatos, Thomas Kuhn, Ernan McMullin and others, a strong consciousness of history is commonplace in philosophy of science – an area usually regarded as one of the strongholds of AP. On the other hand, it is simply not true that all philosophical movements that are usually associated with CP show a particular interest in history that goes beyond the usual backward looking “identity construction”, i. e., historical legitimization of one’s own philosophical project. Such identity construction, though it may include considerable efforts in terms of fencing-off opposite traditions and enclosing relevant historical forerunners, does not necessarily indicate “a strong consciousness of history”. In any case, this holds if “strong” is meant to express a serious interest in history as a self-standing object of philosophical concern and analysis. For instance, such a strong and systematic interest certainly is a driving force behind Heidegger’s philosophical work. Yet this interest is of a very specific stamp. It must not be understood in terms of an interest in the history of philosophical (or other) ideas which Heidegger refers to as “history” (*Geschichte*), but in terms of a historically unfolded existentialist interpretation of temporality (*Geschichtlichkeit*). It is only the latter that, according to Heidegger, is a philosophical topic or, to be more precise, philosophy’s very origin. (It is, of course, part of the history of the AP/CP divide to quarrel about the precise meaning of “precise” methods of doing philosophy. Among the contested meanings that are proposed on part of CP, as indicated in the present context, are the following: “being faithful to the phenomena” and “authentically responding to the phenomena”.) As distinct from this Heideggerian engagement with *Geschichtlichkeit*, Husserl’s early anti-historicist and his later recourse to so-called intentional history in his *Crisis* both are motivated by a far more restricted theoretical concern for history. The latter still is guided by distinctions (e. g. between static and genetic modes of analysis) that Heidegger, from a thoroughly critical point of view, would consider representative of Husserl’s life-long struggling for a “philosophy as science”. The upshot of this very roughly sketched comparison of Husserl’s and Heidegger’s stake in history is that merely indicating a more or less strong or conspicuous interest in history cannot do the job of clarifying the AP/CP divide. As soon as we try to specify the respective interest in history we, once again, realize that our expectation to approach a clear-cut demar-

cation between AP and CP is bound to be frustrated. Rather, it turns out that, as far as the philosophical relevance of history is concerned, there is much more disagreement than agreement within the camps of both AP and CP.

Another frequently heard proposal is that the development of AP is deeply influenced by the “linguistic turn” that was prompted by the writings of Frege, Russell and Wittgenstein and that shifted philosophical interest towards concepts and language. According to Michael Dummett, for instance, “what distinguishes analytical philosophy [...] from other schools is the belief, first, that a philosophical account of thought can be attained through a philosophical account of language, and, secondly, that a comprehensive account can only be so attained” (Dummett 1993, p. 4). However, in our view, this proposal is no more convincing than the previous one. Firstly, it is commonly agreed that philosophy of language lost its place at the center of AP during the 1970ies. Although the study of language is, of course, still part of the canon, its leading role has been superseded by philosophy of mind, epistemology and metaphysics. Secondly, several philosophers in the continental tradition took language no less seriously than their pre-Quinean colleagues in AP. Gadamer’s claim that “man’s relation to the world is absolutely and fundamentally verbal in nature” (Gadamer 2004, p. 471) or Derrida’s remark that “there is no outside-text” (Derrida 1976, p. 158) may serve as a case in point.

Yet another suggestion is to look for meta-philosophical differences in order to draw a meaningful distinction between AP and CP. However, this suggestion, too, is fraught with difficulties. As far as AP is concerned, a natural assumption would be to regard the method of *analysis* as basic for philosophical inquiry. But this provokes a number of obvious follow-up questions: What, precisely, is analysis? And what is it that ought to be analyzed? Is the analysandum of AP linguistic in nature? Without doubt, many (though not all) proponents of AP in the first half of the 20<sup>th</sup> century would have answered in the affirmative. But as indicated earlier, it is no secret that the analysis of language (ideal or ordinary) has lost its supremacy in AP. Conceptual or linguistic analysis has been supplanted by, for instance, “the method of *thought experiments* or, as it might also be called, the appeal to *conceivability* arguments” (Massey 1991, p. 287). More recently, proponents of the so-called “X-phi”-movement even went one step further, urging analytic philosophers to “employ methods that are better suited to the careful study of philosophical intuitions, namely, the methods of the social and cognitive sciences” (Alexander 2012, p. 28). As one would expect, the situation is no less complicated in CP. Here too, one is faced with a plethora of different methodological approaches, ranging from intentional analysis over deconstruction and hermeneutical interpretation to different varieties of a priori and/or transcendental reasoning.

Where does all this leave us? Is the point of the previous remarks that there is nothing to say about the AP/CP distinction except that it is still a powerful institutional, sociological and psychological factor in professional philosophy? Not quite. For one thing, it would be somewhat incongruous to assemble an anthology on a topic that the editors believe to be unworthy of philosophical attention from the very start. Secondly, we do not deny that attempts to chart the contemporary philosophical landscape will have to acknowledge real divergences in philosophical outlook and method – divergences that are not wholly reducible to sociological factors alone. Yet, what we are skeptical of is the idea that issues pertaining to the AP/CP distinction can be fruitfully discussed *in abstracto*, i.e. independently from the actual contents over which philosophers working in different frameworks disagree. On our view, a more productive way to bridge existing rifts in contemporary philosophy is to identify concrete topics of shared interest and to reconstruct the ways in which these topics are introduced and analyzed within different philosophical frameworks. Of course, it would be naïve to assume that this approach is capable of doing away with all disagreements and disputes that separate modern philosophy. This, however, is not the point. Our hope is rather that working through concrete problems will lead to a more realistic and less ideologically biased assessment of the AP/CP divide.

In general, talk about different frameworks may turn out to be useful and protective by warding off too naïve a comparison. Take for instance the issue of intentionality which, at first glance, seems to represent a topic that is perfectly suited to bridge the gap and reunite the opposing parties. However, the mere fact that we find core groups of philosophers on both sides of the AP/CP divide who are concerned with this topic does not indicate substantial agreement on the understanding of intentionality. Let us pick out one crucial issue. While advocates of AP tend to deal with issues of intentionality and phenomenality as if they marked two neatly separable problems of a philosophy of mind, phenomenologists usually argue that the variety of intentional modes of being directed at different things and states of affairs (or different aspects thereof, respectively) is intrinsically connected with varying modes of phenomenality. Given this basic difference, one may broach the issue whether intentionality, typically conceived of in AP style, and intentionality, typically conceived of phenomenologically, really does represent the same phenomenon. Contrary to the opponent's best intentions and (largely or mostly?) unnoticed by them, "dialogue" und mutual approachment may easily turn into talking past each other.

Similar objections suggest themselves if we, for instance, take note of the gap that opens up on occasion of discussions on truth. In AP it has been considered natural to stick to a propositional framework when dealing with truth. Relevant parts of CP have resisted to follow this prevailing approach. As Heidegger has shown in his

famous § 44 of *Being and Time*, it is a core interest of his specific brand of a phenomenological ontology to make explicit those presuppositions that one must already have acknowledged in order to ponder truth exclusively in terms of the doctrine of propositionality. Hence the obvious fact that advocates of AP and CP use a partly overlapping philosophical terminology (e.g. referring to “consciousness”, “intentionality”, “phenomenality”, or “truth”, “evidence”) does not at all convey substantial agreement on how to take a grip on the problems at issue and how to figure out their nature and impact. There is no easy way of understanding relevant differences between AP and CP. Among others, one has to be cautious not to overlook deep differences lurking behind a terminological surface that seems to partly coincide though this may turn out to be entirely illusive. On the other hand, one must not ignore or deny similar methodological interests or similar philosophical theses (e.g. Frege’s and Husserl’s objections against logical psychologism) even if they do not fit into the same overall philosophical projects. The gist of the above is that there is ample room for misconceiving the relation between AP and CP because (dis)agreement does not come wholesale. Entering a more detailed comparison and distinguishing different types and different levels of (dis)agreement it therefore does make sense to argue that we both risk to overlook deep disagreements and to falsely assume total disagreement. In our understanding, presumptions about the so-called gap between AP and CP are less interesting if they are uttered at the level of a general idea of doing philosophy. What is much more important and informative is to dig into concrete philosophical stuff and to deliver fine-grained comparisons that bring to light similarities and differences, and do so with regard to particular topics.

We appreciate that virtually all contributors who devote their papers to the general conference topic have followed our lead to address the AP/CP-distinction by focussing on concrete problems rather than by discussing it in the abstract. We are aware of the fact that this way of approaching the issue may not be the most straightforward one. We believe, however, that opting for a less direct route is sometimes necessary in order to prevent oversimplifications, sterile dichotomies and ideological entrenchments.

## Bibliography

- Alexander, Joshua (2012): *Experimental Philosophy. An Introduction*. Cambridge: Polity Press.
- Ayer, A.J. (1977): *Part of my Life*. Oxford: Oxford University Press.
- Critchley, Simon (2001): *Continental Philosophy. A Very Short Introduction*. Oxford: Oxford University Press.
- Derrida, Jacques (1976): *Of Grammatology*. Baltimore: The Johns Hopkins University Press.
- Dummett, Michael (1993): *Origins of Analytic Philosophy*. Cambridge: Harvard University Press.

- Gadamer, Hans-Georg (2004): *Truth and Method*. London/New York: Continuum.
- Hanson, Russell Norwood (1962): The Irrelevance of History of Science to Philosophy of Science. In: *The Journal of Philosophy* 59/21, pp. 574–586.
- Massey, Gerald J. (1991): “Backdoor Analyticity”. In: Tamara Horowitz & Gerald Massay (eds.): *Thought Experiments in Science and Philosophy*. Savage: Rowman & Littlefield, pp. 261–272.
- Prado, C. G. (ed.) (2003): *A House Divided: Comparing Analytic and Continental Philosophy*. Amherst: Humanity Books.
- Quinton, Anthony (2005): Continental Philosophy. In: Ted Honderich (ed.): *The Oxford Companion to Philosophy*. Oxford: Oxford University Press, pp. 161–163.







Wittgenstein



Peter Hacker

## Can You Have My Pain?

### **Protagonists:**

Gottlob Frege

John Searle

Ludwig Wittgenstein

Peter Strawson

Wolfgang Künne

The setting is an Oxbridge Common Room in Elysium, where the shades of the dead and of the living are conversing after dinner. The French window is open to a great lawn, and the stars are just beginning to show in the twilight. All except Wittgenstein are seated in comfortable, dark brown leather armchairs, around a low table on which their post-prandial drinks are placed. Wittgenstein is seated on a wooden chair, drinking a glass of water. Frege is smoking a pipe. There are some books on the table, some of them open.

**Wittgenstein:** You know, philosophical problems really are like knots we tie in our understanding. The smaller the knot, the harder it is to unravel. If you tug on the thread, the knot gets smaller and smaller – and it becomes the very devil to untie.

**Frege:** Ja, Wittgenstein; we all know that you spin lovely metaphors and similes. What exactly do you have in mind?

**Wittgenstein:** Well, actually, I was thinking of something you wrote?

**Frege:** So! And what displeased you?

**Wittgenstein:** Oh no – what you wrote did not *displease* me. In fact, it intrigued me. Language lays the same traps for us all, but we do not all fall into them. If you, of all people, fell into this one, then it must be very hard to avoid. In your paper, “Der Gedanke” (“Thoughts”), which, you may remember, I didn’t like at the time, you wrote:

Even an unphilosophical man soon finds it necessary to recognize an inner world distinct from the outer world, a world of sense impressions, of creations of his imagination, of sensations, of feelings and moods, a world of inclination, wishes and decisions.

and you went on to call the inhabitants of this “world” of yours “ideas”.

**Frege:** Ach, Wittgenstein – it is not just *my* world. Everyone has such a world. It is the world of consciousness. And what this world contains are *objects* that are of a mental nature. In this I am in agreement with your John Mill.

[*He leans forward and picks up an open copy of Mill's An Examination of Sir William Hamilton's Philosophy from the table*]

Here is what he wrote

I observe that there is a great multitude of other bodies, closely resembling in their sensible properties [...] this particular one, but whose modifications do not call up, as those of my body do, *a world of sensations in my consciousness*. Since they do not do so in my consciousness, I infer they do it out of my consciousness, and that to each of them belongs *a world of consciousness* of its own, to which it stands in the same relation in which what I call my body stands to mine, [...]. Each of these bodies exhibits to my sense a set of phenomena (composed of acts and other manifestations) such as I know in my own case, to be effects of consciousness, and such as might be looked for if each of the bodies has really in connection with it a world of consciousness.

So Mill. Now surely, what he says is undeniable! I disagree with John Mill about the nature of logic and of number, but I am in complete agreement with what he says here.

**Searle:** Yeah! I think that John Stuart Mill and Frege here are right. But I think that we can now put it more precisely. I don't much like this talk about *worlds* – I think that there is only *one* world. As Virginia Woolf said “One of the damned things is quite enough”. But there are things in the world belonging to different ontological categories. There are objective things, and there are subjective things. And by this I don't mean epistemic modes – as when we distinguish between subjective opinion and objective fact. What I mean are *ontological categories*. Now what Frege calls “ideas” are in fact *conscious experiences* or *conscious states*. What I am talking about, and what you [*he turns to Frege*] and Mill were talking about, is *ontological subjectivity*. What characterizes all conscious phenomena is subjectivity. Consider, for example, the statement, “I now have a pain in my lower back”. That statement is completely objective in the sense that it is made true by the existence of an actual fact and is not dependent on any stance, attitudes, or opinions of observers. However, the phenomenon itself, the actual pain itself, has a subjective mode of existence.

**Künne:** John, I really don't understand what you mean by “subjective mode of existence”.

**Searle:** OK, let me put it differently. We're talking about conscious states. All conscious states have a qualitative character that is experienced by the person that has them – by the subject. Conscious states are ontologically subjective. Consciousness has a first-person ontology. It exists only as experienced by a human or an animal subject and in that sense exists only from a first-person point of view.

**Wittgenstein** (*indignantly*): A first-person point of view? What is a first-person point of view? When you have a pain, do you have it from a point of view? I know what it is to express a judgement from a moral, or a political, or an economic point of view – or from my own point of view, when I give my opinion. But what it is to have a pain from a point of view!!

**Strawson:** (*trying to calm him down*) Wittgenstein, I sympathize with your objection to the phrase – but it really doesn't matter, and it will divert us from our theme – which is what Searle is calling “ontological subjectivity”. What is your reaction to what Searle said, Frege?

**Frege:** I think that what Searle is saying is very similar to what I wrote in my old paper that Wittgenstein didn't like. Let me put it *so*: physical objects are actual and objective. Abstract objects are no less objective – they are real, but non-actual. But objects in the inner world are quite different. They are subjective – as Searle says. I wrote, “It seems absurd to us that a pain, a mood, a wish, should go around the world without an owner, independently. A sensation is not possible without a sentient being. The inner world presupposes somebody whose inner world it is.” Maybe Searle is right that talk of worlds here is misleading. What I meant was that, for example, one cannot find pains in the room – only people in pain. One cannot find hopes or fears in the cupboard – only cups; or perceptual experiences in the street – one can only *have them* in the street. What I meant by “inner objects” are conscious experiences that essentially *belong* to someone. Ideas or conscious experiences are *had*. So, for example, one does not perceive one's perceptual experiences – one *has* them. *What* one perceives are what one's experiences are experiences *of*, but one *has* the experiences. Ideas or experiences are *essentially owned*. Things of the outer, physical world, are essentially independent.

**Searle:** Yeah, that's what I meant when I said that subjectivity is an *ontological* category. Experiences have a subjective mode of existence. They exist only when they are experiences had by a human or animal subject. In this respect they dif-

fer from nearly all of the rest of the universe, such as mountains, molecules, and tectonic plates, which have an objective mode of existence.

**Frege:** So. Now, such inner objects are not only owned, they have only *one* owner and they cannot change ownership. I cannot transfer my ideas to you. One cannot bring together in one consciousness a sense-impression belonging to one consciousness and another sense-impression belonging to another consciousness. An object within one inner world – such as my pain – can no more enter another inner world than it can escape to the outer world.

Now, just as mental objects need an owner, so too, they cannot have more than one owner and cannot change owners. That is why I wrote “Nobody else has my pain. Someone may have sympathy with me, [*He chuckles*] but still my pain belongs to me and his sympathy to him. He has not got my pain, and I have not got his feeling of sympathy.” [*He laughs*] Not only can’t you have my pain when I have it, you can’t have my pain even when I cease to have it! If I cease to have a pain, then it ceases to exist – the pain has not only passed, it has passed away. [*He chuckles again; the others laugh. Wittgenstein isn’t amused.*]

**Wittgenstein:** Frege, do you really not see any difficulty here? Why *can’t* I give you my pain? Is it because it is so difficult? Could I *try* to give you my pain?

**Frege:** *Ach*, Wittgenstein – can’t you see that it is not possible. You can kick yourself in the shins and then kick me in the shins – but your pain will be your pain and my pain will be mine.

**Wittgenstein:** You mean that it is “metaphysically impossible” to give you my pain?

**Searle:** Well, I don’t know about metaphysically impossible – but it is certainly ontologically impossible. It is just a plain fact that pain has a subjective mode of existence. Now, what does this mean? It means that because of its subjective mode of existence, its existence is a first-person existence. It must be *somebody’s* pain; and this in a much stronger sense than the sense in which a leg must be somebody’s leg, for example. Leg transplants are possible; in that sense pain transplants aren’t. This is a general truth about conscious states. I have a special relation to my conscious states, which is not like my relation to other people’s conscious states. And they in turn have a special relation to their conscious states, which is not like my relation to their conscious states.

**Frege:** *Ja, Ja*. That’s right.

**Wittgenstein:** No, no. It's not right at all! It is not as if we could *conceive* of unowned pains, only we don't come across any. Nor is it as if we can't come across unowned pains because of some limitation in our constitution. We don't even know what it would be like to come across a pain that wasn't someone's pain. If someone were to say "I found a pain under my bed yesterday", we should have no idea what he meant. What Searle calls "ontological impossibility" is *not a possibility that is impossible*. So both of you have to explain what this impossibility is. After all, it is not a brute fact that could be otherwise – it seems to be a super-fact, a *meta-physical* fact. And neither of you have explained what that means.

**Strawson:** Well, I agree that invoking metaphysical or ontological possibilities and impossibilities is unhelpfully mystifying. It is not something I go in for at all. I think that Frege and Searle have not really got to the bottom of the matter. For the issue is not an ontological or metaphysical one. It is, in a generous sense of the term, a *logical* one. It is a matter of the identification and re-identification of particulars.

It is not helpful to say that it is *impossible* for there to be an unowned pain, or that it is *impossible* for me to have Frege's headache – for all that does is to leave us bewildered by the nature of these putative impossibilities. The truth of the matter is that it *makes no sense* to speak of pains that are no one's pains, or to suggest that the identical pain which was in fact one's own might have been another's. We are not up against adamantine impossibilities. Rather, we have arrived at the bounds of sense. I hope you will indulge me a little and allow me read to you what I wrote on the matter years ago in my book *Individuals*, for I don't think I could put it better today.

*[He leans forward and picks up a copy of Individuals from the table and pages through it until he finds the passage he is looking for]*

Yes, here it is on page 97. This is what I wrote:

We do not in fact have to seek far in order to understand the place of this logically non-transferable ownership in our general scheme of thought. For if we think [...] of the requirements of identifying reference in speech to *particular* states of consciousness, or private experiences, we see that such particulars cannot be identifyingly referred to except as the states or experiences *of* some identified *person*. States, or experiences, one might say, owe their identity as particulars to the identity of the person whose states or experiences they are. From this it follows immediately that if they can be identified as particular states or experiences at all, they must be possessed or ascribable [...] in such a way that it is logically impossible that a particular state or experience in fact possessed by someone should have been possessed by someone else. The requirements of identity rule out the logical transferability of ownership.



It seemed to me then, and it seems to me now, that this *logical* account renders it perfectly intelligible that we should say that there cannot be unowned pains, and that another person cannot have my pain, while avoiding the dubious intelligibility of appealing to metaphysical or ontological impossibilities. The apparent impossibility is actually a limitation of *sense* which is not an arbitrary convention, but has its rationale in the requirements of reference and identification. Pains, and experiences in general, are identifiability-dependent upon the persons whose pains and experiences they are.

[*He lights a cigarette, inhales slightly nervously, and looks at Wittgenstein*]

**Wittgenstein:** I see. [*Pause. He puts his face in his hands.* ]

We must think this through. [*Another pause*]

Now, Strawson has nicely transformed what appeared to be a meta-physical mystery into a logical, or as I would prefer to put it, a grammatical, claim

**Strawson:** [*sotto voce*] I wouldn't. I wouldn't.

**Wittgenstein:** [*continues, without paying attention to Strawson's muttering*] ... a grammatical claim of identity-dependence of pains on owners. He thinks that the identity of pain turns on *whose* pain it is, and therefore another person *can't* have my pain. I may have a headache. If you also have a headache, it seems, we *call* what you have "another headache", and we do not say "You have my headache". This is because the identity of the particular that is your pain depends upon your identity. So I can't have your pain. Now, trying to getting rid of meta-physical necessities is no doubt admirable. But have we put the question marks deep enough down?

**Künne:** Can't Strawson's argument be strengthened by reflecting on pain-location? It is clearly constitutive of physical pain to have a bodily location. It always makes sense to ask "Where is the pain?" or "Where does it hurt?" – and many pains are differentiated by their bodily location. One can't have a headache in one's back or a backache in one's head. The location of pain is obviously *one* criterion of identity of pain. So, if Jack has a toothache and Jill has a backache, then they certainly have different pains, since a backache is a different pain from a toothache. Now, if you grant this much, then you must surely grant that two people can't have the same pain. For Jack's pains are the pains in *his* body and Jill's pains are the pains in *her* body. And Jack's body is different from Jill's body. If there is a pain in Jack's knee, it is Jack's pain. Jill may have a similar pain in her knee. But since her knee is not Jack's knee, the pains are in different places. So it is obvious that two people can't have the identical pain. Difference

of location surely implies numerical difference of pains. Their pains may be qualitatively identical, but they can't be numerically identical. They are, as Strawson says, different particulars. They are owned by different people and they have different locations.

**Wittgenstein:** Hmm. Philosophers should greet each other with the words “Take it slowly”. We must slow down. [*He takes a sip of water and pauses for thought*]

We distinguish, with respect to things like chairs, between the same and being exactly alike. This chair [*he thumps the chair he is sitting on*] is the very same chair I sat on last night when we were talking about logical connectives. It hasn't even been moved from its place. But if it needs repair, it can be replaced with another chair from the same set that is exactly alike. Now, what makes it possible to draw this distinction in this kind of case?

**Strawson:** Well, it is surely this: material objects consist of matter. The chair you are sitting on consists, let us suppose, of fifteen pounds of mahogany. The other chairs of this set likewise consist of fifteen pounds of mahogany. But of course, the wood from which your chair is made is a distinct *specific quantity* from the specific quantities of wood from which the other chairs are made. Two chairs are material space-occupants of the same general kind, but they consist of different specific quantities of matter that cannot occupy the same place at the same time. But, of course, the chairs may, to all intents and purposes, share all their other properties. So they are numerically distinct, but qualitatively identical.

**Wittgenstein:** Hmm. That is on the right lines, although the jargon may be misleading.

We do distinguish, when we speak of things like chairs, between being the same and being exactly alike. If you wish, you may dress that up in the jargon of numerical identity as opposed to qualitative identity. But that can be misleading.

**Strawson:** Why so? I don't see anything misleading here.

**Wittgenstein:** Look! Replacing the expression “the same” by “identical” or “numerically identical” is a typical expedient in philosophy. It makes it look as if we were talking about fine shades of meaning, and are just looking for the right words to hit on the correct nuance. But that is not what is going on here at all. What is going on here is the transposition of an element of one language game into another, where it has no place – like introducing a knight piece into draughts.

It is much easier to tempt ourselves that the headache that two different people have is qualitatively the same but numerically distinct than it is to persuade

ourselves that they are exactly similar but not the same. We are perfectly familiar with the distinction between being the same and being exactly similar in the domain of things like chairs. We are not familiar with its application to experiences in general or pains in particular. We know what it is for two people to have the same headache, but if we are told that A's headache is *exactly similar* to B's, but nevertheless different – we would not understand what we were being told. If your headache and my headache really *are* exactly similar, we should naturally respond, then there is no difference between them.

**Künne:** Of course, pains are not material things. It is people's bodies that are material things. And two people cannot occupy the same space at the same time, just as two chairs cannot occupy the same place at the same time. Now, isn't there at any rate an analogy with pains? You are sitting over there, and I am sitting here. Suppose you have a pain in your knee, and I have a similar pain in my knee. Then you have a pain in a different place from me! The pains may be exactly alike, that is, to use the jargon you dislike, they may be qualitatively identical, but they are obviously not the same – they are numerically distinct, since your pain is in your knee and mine is in my knee.

**Wittgenstein:** All right. Let's not quarrel over the jargon. First of all, reflect on how we determine the location of another person's pain? It is clear that the criterion for where a person has a pain is where he *says* it hurts, where he *points* when he's asked where it hurts, and what part of his body he *assuages*. If two people say that their knee hurts, if they both point at their respective knees when asked where it hurts, or if they both rub their respective knees to assuage their pain, then each has a pain in the knee. And now, isn't this what it is for different people to have a pain in the same place? Isn't this exactly what we *call* "having a pain in the same place"? So isn't wrong to say that different people cannot have a pain in the same place because their bodies are in different places?

**Frege:** Now, Wittgenstein, it is you who is going too fast. [*He relights his pipe and puffs on it*] All your argument shows is that in *ordinary language* the phrase "pain in the *same* location" means the same as "pain in the *corresponding* location". But as I wrote many years ago, "Someone who wants to learn logic from language is like an adult who wants to learn how to think from a child. When men created language, they were at a stage of childish pictorial thinking. Languages are not made so as to match logic's ruler." A *corresponding* location is NOT the *same* location!

**Wittgenstein:** Well, I disagree with your attitude towards ordinary language *and* towards logic. But we don't want to go down that road now. Let me show you that the business of location is not the real issue here. There is an easy way to sidestep the matter of location. Suppose two Siamese twins are joined at the knee. If both have a pain at the point of juncture, then they both have a pain in the knee. But *now* it is *the very same knee* to which they point when asked where it hurts. Here one cannot say that twin A can't have the same pain as twin B, because A's pain is in his knee and B's pain is in his knee, for they have the very same knee. So if they describe the pain in exactly the same terms, must you not admit that they have the same pain?

**Frege:** No! – I don't see that at all. A cannot have B's pain, and B cannot have A's pain. That's the end of the matter. *Another person's pain is another pain!*

**Wittgenstein:** Good. Very good. [*He pauses for thought, hand on brow*]

What your response shows is that the question of pain-location is really a red-herring. The decisive point that moves one to deny that different people can have exactly the same pain has nothing to do with the location of pain. It is because, as one insists *Mine is mine, and yours is yours*. What determines the identity of a pain is above all the *person*. Having a pain, Strawson suggested, is a matter of *logically non-transferable ownership*. So two people cannot have the identical pain, only a similar one. Is that not what you want to argue, Strawson?

**Strawson:** Yes. That is indeed what I thought, and I am still inclined to think it now. You can't have the same pain as I – the numerically identical pain – because you can't own *my* pain. But, of course, you can have the qualitatively identical pain.

**Wittgenstein:** Yes. [*He gets up to pour himself another glass of water from the jug on the sideboard, takes a sip or two, and stands still, thinking for a few moments.*] You all seem to agree on three points. First, that to have a pain is to stand in a relation to what Strawson calls a "particular". I won't say anything about the idea of a particular, Strawson, although I have grave qualms about it. Secondly, that this relation is a relation of ownership. And that thirdly, this kind of ownership is logically non-transferable. Now, ... where should we begin? ... [*He pauses, and then sits down again*]

Let's first enquire into the notion of "logical ownership". Ordinary ownership is a *legal relation* between a person and the thing owned. The car I have belongs to me – it's mine. But I might sell it to another. Then it will belong to him – it will be his.

So ownership is transferable. I might share my car with another. Then we would both own – both *have* – the very same car. So ownership is shareable.

Now, how do we come by the idea of *logical* ownership of pain, or, more generally, of experience? We think that *having* a pain signifies a logical relation of ownership between the sufferer and his pain, just as *having* a car signifies a legal relation of ownership between an owner and a chattel. Similarly, we think that the car's *being mine* signifies a *legal* relation of possession, and so too, the pain's *being mine* signifies a *logical* relation of possession.

[*He pauses, holding his head in his hands*]

Look! We might say that possession is the *representational form* of experience (and hence too of pain). This is how we *present* pain to ourselves in the grammar of our language, just as we present *meaning something by what we say* in the form of an act, and we present *thinking* in the form of an activity. But surface grammar here is deeply misleading. It is remarkable how the auxiliary verb “to have” can lead us astray. Someone's having chattels is indeed a relation of ownership between a person and his chattels. But we also speak of *having* a father and siblings, and we speak of “my father” and of “my sisters”. This does not signify any relation of ownership, but filial and fraternal relationships. We speak of *having* a promissory obligation. That too is a relation, but not a relation of ownership. It is not a relation between us and what we have, namely an obligation. It is rather a normative relation between us and the person to whom we made the promise. We speak of *having* in hundreds of other different kinds of case in which *having* does not signify ownership of *any* kind. And in many cases, such as having a sharp tongue, having a sense of humour, having a good mind, having a good time, having a train to catch, not only is no *ownership* in question, no *relation* of any kind is in question either.

**Searle:** But Professor Wittgenstein, can't you see that each one of us, each subject of experience, stands in a special relationship to *his* or *her* experience. OK, you convince me that talk of ownership here may be misleading. But still, the relation of every person to their own experience is unique. Now I don't mean that each person has what used to be called *privileged access* to their own experience and only to their own experience. I agree that this is just another can of worms. And I don't mean that each person can *introspect* their own and only their own experiences. That's just another confused metaphor. What I mean is that experiences, and consciousness in general, are essentially subjective. They have a first-person ontology. They exist only as experienced by a human or animal subject.

**Wittgenstein:** Well, if you mean that sentient beings are sometimes in pain, that when they cut themselves in the hand, their hand hurts – and that trees and stones are not sentient beings, and so there is no such thing as a tree’s hurting itself or a stone’s being in pain, then of course that is correct. If you mean that when I injure myself, *I* am in pain and you are not, then that too is normally correct. But that does not imply that having a pain is a *relation* between a person and a pain.

To have a headache is for your head to ache – but is that a relation? Between your head and its aching? If you cut your hand, your hand bleeds. Is that a relation between your hand and bleeding?

**Searle:** But what about conscious experience in general. Isn’t it obvious that each person stands in a special relation to his own conscious experiences? The person who has an experience feels the qualitative character of that very experience – and no one else does. That is the subjective character of experience.

**Wittgenstein:** We shall have to talk more about the qualitative character of experience. But, honestly, not now. Look, to have what you are calling a “conscious experience” is not to stand in a relation to something called an experience. To experience fear, or joy, or anger is to be frightened, joyous, or angry. To have a visual experience is to see – but it is not to *stand in a relation* to seeing. Of course, it may also be pleasant or unpleasant – but to enjoy looking at something is not to stand in any relation to looking. But let’s leave all these side-roads for some other evening, and get back to our high-road.

To have a pain is not to stand in a relation to a pain. Pain is not a *relatum* – it is not a kind of *object* at all. That is why I wrote in *my Investigations* that pain is not a *something*, although, of course, it is not a *nothing* either. To have a pain is not to own anything.

**Strawson:** [*lighting another cigarette*] I can see that the terminology of ownership is perhaps misleading. But it seems to me that I can abandon this unfortunate turn of phrase, and yet still continue to argue that different people cannot have the same pain – the numerically identical pain – since pains are identity-dependent upon the person who is suffering. We can only refer to a particular pain in so far as we can identify the person who has the pain. We can only identify a particular pain by reference to who has it.

**Wittgenstein:** Let’s pause a moment and see where we’ve arrived. We know from the Siamese twins that the question of pain-location is really not the crucial issue. We’ve cast doubt on the idea that having a pain is a relation between a

sufferer and a pain. We've repudiated the thought that the subject of pain is the owner of pain. Good. [*He pauses*]

Now, what remains of the claim that my pain must differ from yours – that we cannot have the very same pain? If my pain tallies with yours in phenomenal qualities, such as burning, throbbing, stinging, nagging, if it has the same intensity, and has a corresponding location, what differentiates my pain from yours?

**Frege:** Wittgenstein, you are just going round in circles. You have come back to the original point I made: You can't have my pain, because mine is mine and yours is yours!

**Strawson:** Or to put it in the logical mode: pain is identifiability-dependent upon the person who has the pain. That is why different people can of course have an exactly similar pain – that is, a qualitatively identical pain – but not a numerically identical pain.

**Wittgenstein:** Yes, that is the heart of the matter. What this amounts to is that the property that differentiates my pain from yours is *the property of being mine*. That is, even if there is no difference between my pain and your pain in respect of location, intensity and phenomenal characteristics, still it is a different pain because your is yours and mine is mine. Someone might even say that by Leibniz's Law that makes it a different pain. But this, in effect, is to treat the subject of pain as *the differentiating property of the pain*. But that is absurd. *I* am not a property of my pain. Nor is *being mine*. A substance is not an identifying property of its properties. Nor is *belonging to a given substance* an identifying property of its properties. Let me explain.

*Belonging to me* can be said to be a property of my car, precisely because legal ownership is a genuine relation. *Belonging to me* is a relational property of my car, which the car now has, and will lose when I sell it. But, as we have seen, *belonging to me* is not a property of my pain, precisely because *having a pain* is not a relation and a pain is not a relatum.

Look: to say that *being mine* or *being yours* is a relational property of a pain is like saying that an identifying property of the brown colour of your armchair is that it *belongs* to your armchair. No one would want to say that – for then you would have to say that your armchair cannot have the same colour as my armchair. One can't argue that two different objects can't have the same colour because the colour of A belongs to A and the colour of B belongs to B. That is absurd, because it treats the chair – or "belonging to the chair" – as a differentiating property of its colour.

**Searle:** Yeah, why not? The colours are different *tokens* of the same colour. The token colour of this chair is a different token from the colour of that chair.

**Wittgenstein:** No, no. Perhaps we can accept Peirce's distinction between type and token *words* – although it is not without its problems. But how could we distinguish type and token colours? If there are three blue patches on the wall and two red patches – there are five different coloured patches, but only two different colours, not five. Types and tokens are just a red-herring.

Now, if two objects have a colour of the same hue, intensity and saturation, then they have the same colour. The distinction between being identical and being exactly similar but not the same has no application to colours – or for that matter to weights or lengths.

**Strawson:** Wittgenstein, I don't quite see. And I can't see that you have shown that I was mistaken to say that pains are identifiability-dependent on the subject of pain. Surely, we cannot identifyingly refer to particular experiences, such as pains, or visual experiences, or emotions, except as the states or experiences of some identified person.

**Wittgenstein:** Well, of course I can refer to *your* headache and then go on to say that it is getting worse. But do you mean that the phrase 'my pain' or 'your pain' *identifies* a pain?

**Strawson:** Quite so.

**Wittgenstein:** But not at all! Does the phrase 'the colour of Smith's chair' identify a colour? When I tell you that I want to buy cushions that are the same colour as Smith's chair, do you know what colour the chair is? The phrase "The colour of Smith's chair" does not identify a colour at all – and no more does the phrase "my pain" identify a pain. "My pains" – what pains are they? What counts as a criterion of identity here? – Certainly not *being mine!* If I tell you I have a pain, you still have no idea *what* pain I have, just as if I tell you that I have a coloured glass goblet, you do not know what colour the goblet is. You don't "*identifyingly* refer to a colour" by the phrase "its colour".

**Strawson:** But surely there is a difference. Material particulars can be identified and referred to without reference to any other kind of entity. That is why I wrote that material objects are basic particulars in our conceptual scheme. By contrast, individual events and processes are dependent particulars – for they can be identifyingly referred to only by reference to the material things – the substances



– that are undergoing change. The event or process of *this* chair changing its colour is identified by reference to the chair whose colour fades. It is a different event or process from *that* chair’s fading – precisely because they are different chairs, and the processes of fading are identified by reference to the two different chairs. Now isn’t it just like that with pains or experiences?

**Wittgenstein:** Why should it be? A pain is a sensation, not an event. You may have a pain in your tooth, but a pain in your tooth is not an event in your tooth. The *occurrence* of a sharp stab of pain in your tooth might be said to be an event – but *the pain* you have is not an event. If your toothache gets worse, it is not an event that gets worse, and a throbbing toothache is not a throbbing event. After all, the colour of your hair is not an event, even though your hair’s changing colour – being dyed for example – is.

**Künne:** Professor Wittgenstein, I think that you are mistaken. Of course, we say such things as “Ann and I have the same pain, a throbbing headache in the temples”. But if this were a true identity statement to the effect that Ann’s pain is my pain (rather than to the effect that the kind of pain Ann has is the kind of pain I have), then her pain could not have begun before mine, it could not get worse without mine getting worse, and an executioner could remove her headache by beheading me. Now, in your *Blue Book* dictation, on pages 54–5, you said that this is no argument. But I don’t think that is much of a counter-argument.

**Wittgenstein:** *Of course it is no argument!* The counter argument is *obvious!* If by ‘a true identity statement’ you mean an identity statement such as ‘This is the same chair as the one you saw in the auction last week – I bought it’, then *of course* ‘I have the same pain as you’ is not what you are calling a true identity statement. For what you mean by that phrase is a statement of what you have been calling *numerical identity*. And to be sure, what I have been urging you all along is that the distinction between being the same and being exactly similar, or, as you and Strawson put it, being numerically and being qualitatively identical cannot intelligibly be applied to pains, or, for that matter, to colours. You are assuming, without any warrant whatsoever, that because “I have the same pain as you” is not a case of numerical identity, therefore it is a case of qualitative identity. But that is precisely to beg the question.

*[He jumps up excitedly]*

Look, Ann and I have the same colour hair. That does not mean that when Ann dyes her hair red, my hair turns red too – it simply means that we *had* the same colour hair and now we *no longer* have the same colour hair. This armchair [*Wittgenstein thumps on the arm of Künne’s armchair*] is the same colour as *that*

armchair [*he thumps Strawson's armchair*] – it is not “numerically” the same colour, nor is it “qualitatively” the same colour (there is no such distinction here). It is just the same colour – for both chairs are dark brown. Similarly, if Ann and I have the same headache, a dull throbbing headache in the temples, and Ann takes an aspirin which puts an end to her headache, *of course* that does not mean that my headache will also stop. It means that we'll cease to have the same headache.

**Künne:** Well, I'm not sure. ... But let me try another tack. You said that to suggest that pains are individuated by their owners is tantamount to the bizarre claim that the person who suffers is a property of the pain she has. But that can't be right. I do not declare the Earth to be a property of its axis by saying that the axis of the Earth is, just as such, different from that of any other heavenly body. I do not declare Socrates to be a property of his death by saying that the death of Socrates is, just as such, different from the death of any other human being.

Actually, I don't think that you are right to compare pains to chairs at all, you should have compared pains to axes of objects or to deaths of human beings.

**Wittgenstein:** Come now, Künne. I contrasted pains with chairs and compared them with *colours*. But I hope everyone realizes that I am *not* suggesting that the grammar of pain is just like the grammar of colour – it most certainly is not. After all, you can be blue all over [*he smiles*] like the ancient Britons, but you can't be pain all over. But there is an important analogy between having a pain and being of a certain colour – an analogy made visible by the similarity of their grammars with respect to sameness and difference. The fact that this chair was painted green later than that chair, does not show that they are different in colour, just as the fact that your headache began later than mine does not show that we do not have the same headache – a mild throbbing pain in the left temples.

**Künne:** Why shouldn't we consider pains to be in the same boat as grins? Jack's grin is identity dependent on him. That's why Lewis Carroll's Cheshire cat is funny – for no cat, no smile. So too a person's pain is identity-dependent on the person.

**Wittgenstein:** Well, Jack's grin lasts as long as Jack is grinning. When he stops grinning he doesn't have a grin on his face any longer. But, of course, Jack's son may have his father's grin. The reason the Cheshire cat amuses us is because there are no grins without grinning faces, no smiles without smiling faces –

not because the grinner or smiler supplies us with a criterion of identity for grins and smiles.

Now let me get back to your previous two analogies – between pains and events and between pains and axes of rotation. I have already explained why pains are not comparable to events. I agree that the death of Socrates is a unique event, but Socrates’s headache is neither unique nor an event – it is a sensation. I suggested that, in order to shed light on the identity of pains, we should compare the grammar of sensation with that of colour. You suggested that we should compare pains to axes of rotation? Well, is it even true that the axis of rotation of the Earth is, as you put it, *just as such*, different from that of any other heavenly body? What is the axis of the Earth? Do you mean the axial tilt? Well, what is it?

**Searle:** As far as I remember, it’s 23 or 24 degrees.

**Wittgenstein:** Just so. And suppose that one of the many moons of Jupiter also has an axial tilt of 23 to 24 degrees. Won’t it then have the same axis as the Earth? [silence]

Or do you mean the *length* of the Earth’s axis?

**Searle:** That’s about 7,900 miles.

**Wittgenstein:** So. There are no other planets in the solar system with *that* axial length – but among the billions of planets scattered throughout the universe, it is more than likely that many will have exactly the same axial length.

Künne, the simple fact is that the phrase “the axis of rotation of the Earth” does not say *what* axis of rotation the Earth has. No criterion of identity is associated with the phrase “the axis of the Earth”. But if you tell us that the axial tilt of the Earth is 23 degrees and its axial length is 7,900 miles, then there is no reason why some other heavenly body should not have exactly the same axis of rotation. And to be sure, the axis of the Earth can and does change – and if it does, it will still have *an* axis – indeed *its* axis, but no longer the same axis. And if it had the same axis as the planet Lauflin, then after its tilt diminishes, it will no longer have the same axis as Lauflin.

The phrase ‘axis of the Earth’ refers to the axis of rotation of the earth, just as the phrase “A’s pain” refers to A’s pain – but neither phrase *identifyingly* refers. If you say “No other heavenly body can have the same axis of rotation as the Earth”, we can ask you what the axis of rotation of the Earth is, what its tilt and length are – and when you answer, you give us criteria of identity for its axis – and, of course, other heavenly bodies may have the same axis. Similarly, Jack’s pain refers to the pain Jack has, but it does not *specify what pain Jack has*.

Once you specify the pain as a throbbing headache in the left temple, then we have a criterion of identity, and another person may well have the same pain.

**Frege:** [*waving his pipe indignantly*] But Wittgenstein, do you really mean to say that you can have my pain? Surely that is nonsense.

**Wittgenstein:** Well, it depends what you mean. If A has eaten some bad food, he might get a headache. Now B may eat the same food, and half an hour later he may have a headache with exactly the same phenomenal features, and say to A: “Now I’ve got your headache”. What he means is that now he has the same headache A had. That is innocuous. It doesn’t mean that A’s headache has ‘migrated’ from A’s head to B’s head. It doesn’t imply that A has ceased to have a headache. If someone in such circumstances were to say ‘I have your headache’, we might well take him to mean that he now has the same headache as you – and that makes perfectly good sense, as long as we don’t conceive of *having a pain* as a relational property between a person and a pain. The fact that A’s head started to hurt before B’s does not mean that A had a different headache than B – only that the event of A’s head starting to ache is a different and prior event from B’s head starting to ache. So too, your armchair may have become brown before Strawson’s did – if the leather on your armchair was dyed before the leather on Strawson’s armchair. But that does not show that the two chairs are not exactly the same colour.

**Strawson:** But Wittgenstein, surely the phrase “I have your headache” grates – and we are inclined to want to exclude it as senseless. We are all inclined to say “You can’t have my headache”.

**Wittgenstein:** Yes. That’s all right. You can exclude this misbegotten phrase from currency. But only on two provisos. First, that this does not imply that we can’t have *the same* headache. Secondly, that you realize that if *you* can’t have my headache, then *I* can’t have my headache either.

**Frege:** *Ach*, Wittgenstein, what on earth do you mean?

**Wittgenstein:** What I mean is that if the phrase “I have” does not have the logical multiplicity to accept “your pain”, then it doesn’t have the logical multiplicity to accept “my pain” either. Isn’t that obvious?

**Frege:** No.

**Wittgenstein:** But it is obvious. Look: what does “my pain” *mean*? It means “the pain I have”. But if so, then “I have my pain” means “I have the pain I have”, which says nothing about my pain (just as “Either it’s raining or it’s not raining” says nothing about the weather).

**Künne:** I’m exhausted

**Strawson:** I think we all are. Let’s have another glass of this excellent fortified nectar, and go out into the garden and look at the dazzling stars.

**Frege:** *Ja*, I need a large schnapps after that!

**Searle:** Wittgenstein, thank you – that was fun. Let’s go and watch the night sky.  
[*The others all pour themselves fresh drinks and wander out into the garden.*]

**Wittgenstein** [*to himself*]: Very interesting. Very interesting. I wonder... [*He follows them through the French windows with his glass of water*]

Martin Kusch

# Wittgenstein's *On Certainty* and Relativism

**Abstract:** One important strand in the contemporary debate over epistemological relativism focuses on the question whether, and to what extent, Wittgenstein in *On Certainty* (1969) leaned towards this position. This paper is a contribution to this strand. My discussion has four parts. I shall begin by outlining my interpretation of Wittgensteinian certainties. Subsequently I shall briefly introduce some central arguments for and against attributing epistemic relativism to *On Certainty*. This will be followed by a sketch of the cluster of ideas that – on my analysis – define important versions of the doctrine in question. And finally I shall give my own interpretation of *On Certainty* in relation to epistemic relativism.

## 1 *On Certainty* – A Brief Reminder

*On Certainty* describes as its central interest “cases in which I rightly say I cannot be making a mistake [...]” And the passage continues: “I can enumerate various typical cases, but not give any common characteristic.” (§ 674) In this section I shall offer a classification of these “various typical cases”. On my count these cases fall into five *epistemic categories*. These categories differ in how certainties relate to evidence, justification and knowledge.

*Category I* consists of beliefs for which we have evidence that is both overwhelming and (at least in good part) dialectically mute. There are eight types of such cases:

- (I.1) Perceptual/proprioceptive beliefs about one's own body (e. g. “[...] here is a hand.” § 1)
- (I.2) Perceptual beliefs about close familiar medium-size objects (e. g. “I believe there is an armchair over there.” § 193)
- (I.3) Introspective beliefs (e. g. “I am in pain.” § 178)
- (I.4) Memory beliefs of salient features of one's autobiography (e. g. “I have lived in this room for weeks past.” § 416)
- (I.5) Beliefs based on simple deductive reasoning; e. g. calculations (e. g. “ $12 \times 12 = 144$ ” § 43)
- (I.6) Simple inductive beliefs, e. g. about familiar simple objects (“After putting a book in a drawer, I assume it is there [...]” § 134)
- (I.7) Testimonial beliefs based on parents' or textbook testimony (“[...] textbooks of experimental physics [...] I trust them.” § 600)
- (I.8) Semantic beliefs (e. g. “My name is L.W.” § 425)

The epistemic status of category I cases – (I.1) to (I.8) – can be summed up as follows. The confidence with which we hold these beliefs is “empirically based” (Wright 2004, p. 36). I have, for instance, a lifelong experience of being called “Martin Kusch”. I also have a lengthy experience of the flat I live in. And I have a protracted experience of me being regarded as calculating reliably and reasoning competently both deductively and inductively. And yet, even though our evidence is overwhelming in these cases, much of it is not accessible. Most of it lies outside our present point of view; most of it is forgotten. And thus it cannot be disclosed and shared. This is the reason why I call it “dialectically mute”: we cannot marshal it against someone who denies our certainties of this category.

*Category II* is the class of mathematical propositions that have “officially been given the stamp of incontestability” (§ 655). As Wittgenstein explains at greater length in the *Lectures on the Foundations of Mathematics*, mathematical propositions are based on the overwhelming empirical evidence of experiments: for example, the experiment showing that *two balls and two balls balances four balls*. We turn such initially empirical propositions into mathematical propositions by immunizing them against empirical refutation (Wittgenstein 1976, p. 98).

*Category III* cases are fundamental empirical-scientific beliefs (e.g. “The earth is round.” § 291; “Water boils at 100 °C.” § 293). Here too we have very strong and tractable evidence for the beliefs. And again we take the strength of the evidence as a reason to go further and immunize the beliefs. Other than in the case of mathematical propositions however, we do so not “once and for all” but only “for the time being”. In other words, our readiness to give up one of these beliefs is greater than with respect to *Category II*. Note that Wittgenstein is willing to say that certainties in *Category III* can be known:

291. We know that the earth is round. We have definitively ascertained that it is round. We shall stick to this opinion, unless our whole way of seeing nature changes.

*Category IV* embodies beliefs that constitute what we might call “domains of knowledge”. I mean certainties like “[...] the earth has existed for many years past” (§ 411), or “[...] the earth exists” (§ 209). These beliefs do not fit the model of “beliefs supported by strong empirical evidence transformed into certainties via immunization”. *Category IV* beliefs are beyond empirical confirmation. Or, put differently, only by presupposing them can we show them to be true.

Finally, *Category V* consists of fundamental religious beliefs, like “Jesus only had a human mother.” (§ 239) As Wittgenstein emphasises in the *Lectures on Religious Belief* of 1939: “Reasons [for religious beliefs] look entirely different from normal reasons. They are, in a way, quite inconclusive.” (Wittgenstein 1966,

p. 56; cf. Kusch 2011) There are reasons and evidence in this realm too, but they are of a different kind from the evidence in the other categories.

## 2 For and against epistemic relativism in *On Certainty*

The debate over the question whether *On Certainty* leans towards epistemic relativism is not new. And thus there already exist a number of arguments pro and contra. I now shall try to give a brief summary of these arguments. I begin with the “pro” case and shall focus on three authors, Paul Boghossian, Anthony Grayling and Rudolf Haller. All three base their argument on specific passages. Grayling and Haller cite the following paragraphs in evidence:

- 65. When language-games change, then there is a change in concepts, and with the concepts the meanings of words change.
- 95. The propositions describing this world-picture might be part of a kind of mythology.
- 99. And the bank of the river consists partly of hard rock, subject to no alteration or only to an imperceptible one, partly of sand, which now in one place now in another gets washed away, or deposited.
- 166. The difficulty is to realize the groundlessness of our believing.
- 256. On the other hand a language game does change with time.
- 336. But what men consider reasonable or unreasonable changes.

§§§ 65, 99, 256, and 336 all emphasize the occurrence of fundamental change: in language-games, concepts, word meaning, and rationality. This for Grayling is “classically strong relativism” since it “constitutes a claim that the framework within which claims to knowledge and challenges of doubt equally make sense is such that its change can reverse what counted as either” (2001, p. 308). §§§ 94, 95, and 166 in turn raise the question “what if the background – e.g. your picture of the world – [were] different?” (Haller 1995, p. 229) Does not Wittgenstein imply that there is nothing that can be said about such a scenario? At least nothing evaluative? It appears that “we remain without any ground for the decision between conflicting judgements based on different world pictures” (Haller 1995, p. 230).

Boghossian suggests that it is first and foremost paragraphs §§§ 609–612 that express a commitment to epistemic relativism (2006, p. 107):

- 609. Suppose we met people who [...] instead of the physicist [...] consult an oracle. [...] – If we call this ‘wrong’ aren’t we using our language-game as a base from which to combat theirs?



610. Are we right or wrong to combat it? Of course there are all sorts of slogans which will be used to support our proceedings.
611. Where two principles clash that cannot be reconciled with one another, then each man declares the other a fool and heretic.
612. I said I would ‘combat’ the other man,—but wouldn’t I give him reasons? Certainly; but [...] at the end of reasons comes persuasion. (Think what happens when missionaries convert natives.)

Boghossian’s take can be supported by the following considerations. § 609 suggests that our criticism of the tribe’s use of an oracle is based on the biased background of our own language-game. § 609 also hints that our negative response invariably has an unduly aggressive character: we are not just stating our view, we are “fighting” (*bekämpfen*) their central beliefs. § 610 leaves it open whether our fight is justified or not. Indeed, the “of course” (*freilich*) even raises the prospect that our attack may be supported by nothing else than “all sorts of slogans”. § 611 might be taken to say that there are no worldview- or system-independent considerations in virtue of which one knowledge system – our or the tribe’s – can be said to be more correct than any other, and that name-calling is the only option. And § 612 seems to capitalize on the limits of reasoning across cultural boundaries. The members of the tribe cannot be rationally convinced, they can only be persuaded or converted.

Authors opting for “contra” also often invoke specific passages in *On Certainty* to make their case. First, Michael Williams argues that Wittgenstein opposes relativism in § 286 of *On Certainty* (Williams 2007, p. 108):

286. We all believe that it isn’t possible to get to the moon; but there might be people who believe that that is possible and that it sometimes happens. We say: these people do not know a lot that we know. And, let them be never so sure of their belief – they are wrong and we know it.

If we compare our system of knowledge with theirs then theirs is evidently the poorer one by far.

Second, Williams also finds Wittgenstein opposing relativism in § 108 where it is claimed that we “should feel ourselves intellectually very distant from someone who said” that “it isn’t possible to get to the moon; but there might be people who believe that it is possible and that is sometimes happens” (Williams 2007, p. 108).

Third, Georg-Henrik von Wright (1982) believes that Wittgenstein rejects, or at least limits, relativism in § 92:

92. [...] Remember that one is sometimes convinced of the correctness of a view [here a world-view; MK] by its simplicity or symmetry, i.e., these are what induce one to go over to this point of view.

This passage seems to treat simplicity and symmetry as values that are valid across worldviews.

Fourth, a number of interpreters hold that Wittgenstein's talk of "persuasion" or "conversion" does not call for a relativistic rendering. These passages (§§§ 92, 612) are no more than reminders that the tribesmen need to learn a lot of physics. Even though, as a matter of fact, our scientific considerations are not recognized by the tribesmen, the latter *should* accept the former (Putnam 1992, p. 176; M. Williams 2007, p. 108).

The fifth argument insists that it is the very way in which Wittgenstein ties meaning to specific language-games that blocks the route to relativism. In order for the members of two cultures to disagree, they must be able to grasp the same thought or conceptual content. After all, relativism is often tied to the idea of "faultless disagreement". But for there to be a disagreement, there has to be a common content to disagree over. For instance, for you and me to disagree (faultlessly) over the taste of rhubarb, we must give conflicting answers to the question 'does rhubarb taste good?' As long as we are within the same culture, there is no problem. Alas, the situation changes once we talk about different cultures. As the second anti-relativistic argument has it, Wittgenstein's thinking on languages and language-games simply does not allow for the identity of conceptual content across cultures. And hence there is no space for cultural relativism (Dilman 2004, p. 176).

Sixth, epistemic relativists assume that there can be, and perhaps even are, fundamentally different epistemic systems or practices. But a number of interpreters of Wittgenstein deny that he allows for the possibility of such fundamental differences. This line goes back to Bernard Williams (1981) and Jonathan Lear (1984). More recently, Annalisa Coliva (2010a, p. 201; 2010b, p. 20) and Duncan Pritchard (2009) have claimed that at least some certainties have to be universal, and that disagreements between cultures can always be rationally resolved on that shared basis. And thus there is no space for epistemic relativism.

### 3 What is epistemic relativism?

In order to adjudicate the debate over epistemic relativism in *On Certainty* we need a more precise rendering of what this "ism" amounts to. In this section I propose such a rendering. I have arrived at this account by collecting definitions and characterisations of relativism from both friends and foes of the view, in-

cluding Barry Barnes and David Bloor (1982), Paul Boghossian (2006), Gilbert Harman (Harman and Thomson 1996), Gideon Rosen (2001), Frederick F. Schmitt (2007), Bernard Williams (1981), and Michael Williams (2007). Wittgenstein's texts were not consulted. My aim was to have an independent and stable standard against which to measure Wittgenstein's position. The characterisation could of course be developed at much greater length than I have space for here.

(1) *Dependence*: A belief has an epistemic status (as epistemically justified or unjustified) only relative to an epistemic system or practice (=SP). (Cf. Williams 2007, p. 94).

I write “epistemic system or practice” (subsequently “SP”) in order to indicate that *Dependence* is compatible with both a “generalist” and “particularist” understanding of epistemology. *Dependence* also allows for a further choice regarding SPs. In saying that a belief has an epistemic status (as justified or unjustified) only relative to an SP, the relativist might refer to either the SP of the relevant believer, or to the SP of the attributor or evaluator. (Cf. White 2007; Williams 2007; Boghossian 2006, p. 72).

(2) *Plurality*: There are, have been, or could be, more than one such epistemic system or practice.

Given *Plurality*, relativism is compatible with the idea that our current SP is without an existing alternative. Moreover, *Plurality* permits the relativist to be highly selective in choosing those SPs with respect to which relativism applies. She might for example restrict her relativistic thesis to just two SPs. For instance, one can be a relativist about science and religion, considering each an SP in the sense of *Dependence*.

(3) *Exclusiveness*: SPs are exclusive of one another. This can take two forms:

(a) *Question-Centered Exclusiveness*: There are sets of yes/no questions to which SPs give opposite answers.

(b) *Practice-Centered Exclusiveness*: There are no yes/no questions to which SPs give opposite answers since their concepts and concerns are too different. SPs exclude each other in that the consequences of one SP include such actions or behaviors as are incompatible with the actions and behaviors that are consequences of other SPs. Users or members of one SP are not able to fully understand the actions and behaviors common in other SPs. (Williams 1981; 1985)

*Exclusiveness* tries to capture the sense in which – under a relativistic conception of their relationship – SPs have to conflict. This idea is in tension with the further assumption, made by some authors, that relativism concerns incommensurable SPs (here such incommensurability involves differences in categories that rule out an identity of propositional content across these SPs). The option of *Practice-Centered Exclusiveness* covers this eventuality. Two SPs can be compared, and can conflict, when they lead to, or require, incompatible forms of action and behavior in an

at least roughly specifiable area of human affairs. The requirement that the area of human affairs be specifiable safeguards that there is a certain degree of comparability. And the demand that the forms of action and behavior involved are incompatible, makes sure that the condition of conflict is met.

(4) *Notional Confrontation*: It is not possible for a group  $G$  holding an epistemic system or practice  $SP_1$ , to go over to an epistemic system or practice  $SP_2$  on the basis of a rational comparison between  $SP_1$  and  $SP_2$ . But  $G$  might be converted to  $SP_2$  without losing its hold on reality. (B. Williams 1981, 1985)

A “notional” confrontation differs from a “real” confrontation; in the case of the latter a rationally motivated ‘switching’ is possible. A conversion is not an altogether irrational event. Being converted to a cause is not the same as being self-deceived, brainwashed or drugged. There is no assumption that a conversion is a phenomenon of psychological or social pathology. This idea is captured by the phrase “without losing its hold on reality” (Williams 1981, p. 139).

(5) *Symmetry*: Epistemic systems and practices must not be ranked.

*Symmetry* can take a number of different forms that are worth distinguishing.

(a) *Methodological Symmetry*: All  $SP$ s are on a par vis-à-vis social-scientific investigations.

The best-known version of *Methodological Symmetry* is perhaps the “Symmetry” or “Equivalence Postulate” of the “Strong Programme” in the “Sociology of Scientific Knowledge”: “... all beliefs are on a par with one another with respect to the causes of their credibility” (Barnes and Bloor 1982, p. 23). I generalize this “postulate” in order to detach it from the requirement that explanations must be causal.

(b) *Non-Neutrality*: There is no neutral way of evaluating different  $SP$ s.

*Non-Neutrality* is the main consideration usually invoked in defense of *Symmetry*. It does not preclude the possibility that some  $SP$ s agree on the standards by which their overall success should be judged. What *Non-Neutrality* denies is that such local agreement justifies the hope for a global or universal agreement.

(c) *Equality*: All  $SP$ s are equally correct.

Most characterizations of relativism – by friends and foes alike – take *Equality* to be the natural consequence of *Non-Neutrality* and thus the best way to spell out *Symmetry*. But *Equality* makes a stronger claim than *Non-Neutrality*. This becomes easy to appreciate once we remember the typical challenge to *Equality*: what is the point of view from which *Equality* is asserted? On the face of it, *Equality* appears to presuppose a neutral point of view from which we can somehow see that all  $SP$ s are equally correct. And this very claim jars with *Non-Neutrality*.

(d) *Non-Appraisal*: For a reflective person the question of appraisal of (at least some other)  $SP$ s does not arise.

*Non-Appraisal* seems to avoid the problems of *Equality*, while capturing the important core of *Non-Neutrality*. It is motivated by the thought of “intellectual

distance”: the idea that a reflective person holding one *SP* might come to the conclusion that her own “vocabulary of appraisal” simply does not get a proper grip on the judgments and actions of another *SP*. It is not that this vocabulary could not possibly be applied at all – it is rather that such application seems forced, artificial and contrived (Williams 1981, pp. 141–142).

*Dependence, Plurality, Exclusiveness, Notional Confrontation* and *Symmetry* are (in some version or other) essential features for relativism. The remaining four elements are not essential, though they are its very frequent, almost regular, bed-fellows. They are: *Contingency, Underdetermination, Groundlessness* and *Tolerance*.

(6) *Contingency*: Which epistemic system or practice a group *G* or individual finds itself holding is a question of historical contingency.

If the history of *G* had been different – for instance, if *G* had encountered certain other groups at certain points, or if *G* had lacked the means to engage in certain types of costly investigations – *G*’s current *SP* would be substantially, perhaps even radically, different from what it is now. The contingency might reach deep: even those beliefs that one ordinarily deems ‘self-evident’ or ‘completely certain’ can be discovered to be contingent. Becoming aware of the contingency of one’s views in this sense can, but need not, undermine the strength of one’s conviction (Rosen 2001).

(7) *Groundlessness*: There can be no non-circular epistemic justification of one’s own epistemic system or practice.

*Groundlessness* is rarely formulated as a distinct ingredient of epistemic relativism. But it is sometimes invoked in arguments meant to establish the truth of relativism. For instance, it is occasionally put forward that epistemic relativism results from the recognition that all *SPs* are on a par insofar as none of them is able to justify itself without moving in an (illegitimate) circle (cf. Williams 2007, p. 95). (Needless to say, it might well be possible to advocate a form of relativism that permits such circularity.)

(8) *Underdetermination*: Epistemic systems and practices are not determined by facts of nature.

*Underdetermination* is not to be confused with the thesis that the world has no causal impact on *SPs* at all. The relativist is not – or need not – be committed to the view that *SPs* are completely arbitrary. His point is rather that (many) more than one *SP* is compatible with the given causal impact of the world.

(9) *Tolerance*: Epistemic systems or practices other than one’s own, must be tolerated.

## 4 Relativistic themes in *On Certainty*

Having introduced above my interpretation of Wittgensteinian certainties, the arguments *pro* and *contra* a relativistic rendering of *On Certainty*, and my characterization of epistemic relativism, I can at last turn to my own assessment of whether Wittgenstein in his last notebooks sympathizes with this position. I take it to be a shortcoming of previous discussions of the issue, that they build a reading of *On Certainty* on the basis of just a few select passages. This jars with the fact that the book collects material for a “well-ordered synopsis” (*übersichtliche Darstellung*) (Wittgenstein 2001, § 122) of our – primarily Wittgenstein’s own – responses to people who deny, or who seem to deny, one or more of our certainties in different circumstances. The point of this well-ordered synopsis is to emphasise the variety of, and patterns in, our responses to such denials. Some of our responses are relativistic, others are not. It is easy to see that, if I am right about this, then it must be a mistake to count the book as a whole as either relativistic or anti-relativistic.

There are about thirty scenarios in *On Certainty* in which someone denies, or seems to deny, a certainty. Each of these requires a more detailed discussion. I offer such details in a book in progress. Here I have to confine myself to a brief summary of some of my central results. In order to do this most economically, I need to introduce the idea of varying cultural distance between Wittgenstein and the certainty-denying people he imagines encountering. The coarse-grained categorisation of distance Wittgenstein seems to work with can be captured in Figure 1.

The centre of the concentric circles is occupied by Wittgenstein himself. In the following rings around him, at increasing distance, are friends (e.g. G.E. Moore) in his own culture, strangers in his own culture, and children. The remaining rings are people outside of Wittgenstein’s culture. Ring **a** is for members of other cultures that Wittgenstein is willing to treat as *virtual members* of his own culture. The outmost ring **c** is for members of other cultures that are too distant to be counted as virtual members of Wittgenstein’s own culture, and that nevertheless are too intelligent to be dismissed. They are not properly appraisable. Finally, there are members of other cultures that have an ambiguous status (Ring **b**). Our interest is with the person in the centre. How does he or she respond to people at varying social-cultural distances that seem to deny what are certainties of various types for the him or her?

I will now give examples of Wittgenstein’s own responses. When a friend – e.g. Moore – straightforwardly denies a certainty then Wittgenstein is inclined to regard him as “demented” or “insane”.

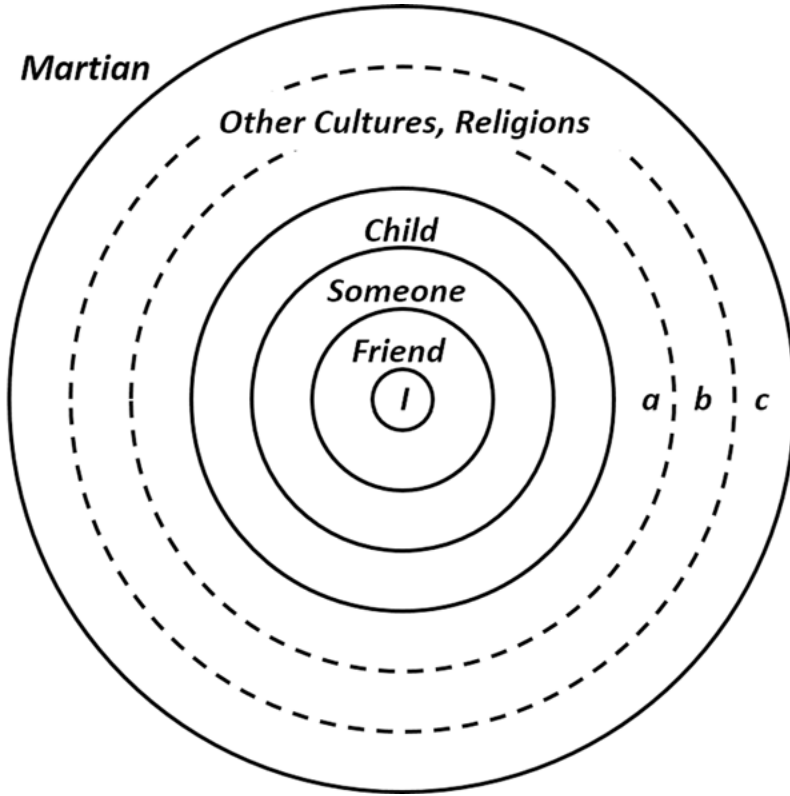


Figure 1

71. If my friend were to imagine one day that he had been living for a long time past in such and such a place, etc. etc. [where I know he has not lived; MK], I should not call this a *mis-*  
*take*, but rather a mental disturbance, perhaps a transient one.
155. [...] If Moore were to pronounce the opposite of those propositions which he declares cer-  
tain, we should not just not share his opinion: we should regard him as demented.

Wittgenstein treats other adult members of his own culture similarly to how he treats friends.

217. If someone supposed that *all* our calculations were uncertain and that we could rely on  
none of them [...] perhaps we would say he was crazy.
257. If someone said to me that he doubted whether he had a body I should take him to be a  
half-wit. [...]

Children receive a more charitable treatment. In their case – at least regarding some categories of certainties – Wittgenstein is willing to offer arguments, explanations, criticism, and education.

310. A pupil and a teacher. The pupil will not let anything be explained to him, for he continually interrupts with doubts, for instance as to the existence of things, the meaning of words, etc. The teacher says “Stop interrupting me and do as I tell you. So far your doubts don't make sense at all”.
322. What if the pupil refused to believe that this mountain had been there beyond human memory?

We should say that he had no *grounds* for this suspicion.

Moving further outwards in the system of concentric circles gives me an occasion to revisit some of the key paragraphs cited as evidence by advocates and opponents of relativistic readings of *On Certainty*.

I agree with Michael Williams that § 286 shows little inclination towards relativism. Wittgenstein clearly thinks that sometimes we are willing – or have no choice but – to dismiss member of other cultures as ignorant and as lacking in knowledge. Thus the tribesmen who in 1950 insist that someone has been to the Moon – and who thereby deny one of our fundamental empirical-scientific beliefs – are “people who do not know a lot that we know”.

Note however that the situation changes in a context in which the tribesmen's statement does not deny one of our scientific certainties. These are contexts in which the statement belongs to the domains of religion or magic. Thus § 92 considers a king who has been told since childhood that “the earth has only existed ... since his own birth.” Wittgenstein likens the king's belief to magical beliefs about one's ability to make rain. This suggests to Elisabeth Anscombe (1976) that the king of § 92 is best thought of as a religious leader such as the Dalai Lama. Wittgenstein imagines Moore trying to convince the king of Moore's certainty that the earth has existed since long before our birth. And he goes on:

92. [...] I do not say that Moore could not convert the king to his view, but it would be a conversion of a special kind; the king would be brought to look at the world in a different way.

Remember that one is sometimes convinced of the *correctness* of a view by its *simplicity* or *symmetry*, i.e., these are what induce one to go over to this point of view. One then simply says something like: “*That's* how it must be.”

What is striking here is the absence of any “they are wrong and we know it” (§ 286). It is also noteworthy that “simplicity” and “symmetry” are not introduced as universal principles adjudicating between systems of belief. They



“sometimes” help to convince us, sometimes they do not. That makes it doubtful whether § 92 is an argument against relativism, as von Wright assumes.

This is confirmed by another passage in which someone like “the Dalai Lama” is at issue again:

238. I might therefore interrogate someone who said that the earth did not exist before his birth, in order to find out which of my convictions he was at odds with. And then it *might* be that he was contradicting my fundamental attitudes, and if that were how it was, I should have to put up with it.

To “put up with it” is to accept that our categories of evaluation do not get a proper grip on this system or practice. Our response is thus one of “Symmetry as Non-appraisal”, in line with Practice-centred Exclusiveness and Notional Confrontation.

Interestingly enough, having convinced himself that this is indeed the proper response to certain disagreements in the realm of magic and religion, Wittgenstein then goes on to draw an analogy between the disagreements between believers and unbelievers on the one hand, and our response to past periods with their different conceptions of the reasonable and the unreasonable, on the other hand:

336. But what men consider reasonable or unreasonable alters. At certain periods men find reasonable what at other periods they found unreasonable. And vice versa.  
But is there no objective character here?

Very intelligent and well-educated people believe in the story of creation in the Bible, while others hold it as proven false, and the grounds of the latter are well known to the former.

Wittgenstein here seems to suggest that “Symmetry as Non-appraisal” might have application not just in our encounters with certain forms of religion and magic, but also in other areas, for instance in our relationship to past forms of science.

Recall that Michael Williams also cites § 108 as a piece of evidence against relativism in *On Certainty*. This is the section where Wittgenstein considers someone who claims to have been to the moon, but who cannot tell us how he managed to get there. Wittgenstein comments: “We should feel ourselves intellectually very distant from them.” I am unconvinced that this comment expresses opposition to relativism. The relativist too can feel “intellectually very distant” from someone who denies one of his certainties. Indeed, in light of the above, we could read the “we should feel ourselves intellectually very distant from them” as an expression of Symmetry as Non-appraisal rather than as a dismissal of relativism. We are dealing with a response based on unfamiliar principles of reasoning and we do not know how to evaluate the response or these principles.

This leaves the longish consideration stretching from § 608 until § 612. I am with Boghossian and others who find strong relativistic sentiments expressed in these paragraphs. I cannot see the strengths of Putnam's and Williams's suggestion according to which Wittgenstein means to insist that the tribesmen need to learn our science. If that is the upshot of these paragraphs then why didn't Wittgenstein say so explicitly? At the same time, I am willing to grant that these five sections do not all point one way: there is a dialogue going on here. One of the voices calls the tribesmen "primitive" for consulting oracles (§ 609) and declares itself willing to "combat" these men (§§§ 609–610). And yet, the conversation terminates in the thought that we can do no better than "persuade" (rather than *convince*) the natives. And this "persuasion" is compared to religious conversion (§ 612).

I also need to briefly comment on the two more general arguments against attributing relativism to *On Certainty*. One was that Wittgenstein cannot be advocating relativism since he denies identity of propositional content across language games (and thus cultures). I have three brief replies. First, it seems to me historically inaccurate to restrict relativism to the idea of a disagreement between two cultures (contexts, epistemic or moral systems) over one and the same propositional content. Many self-proclaimed relativists of the past were relativists in a wider sense that includes Practice-centred Exclusiveness. Second, in *On Certainty* Wittgenstein never emphasises that, say, "the earth has existed since long before my birth" expresses different propositions for him, or Moore and the king (or the Dalai Lama). And third, Wittgenstein is generally not comfortable with dilemmas of the form: either two sentences express the *same* proposition or they express *altogether different* ones. His general thought style in the *Philosophical Investigations* and *On Certainty* is to emphasise degrees and continuities rather than sharp boundaries.

Finally consider the argument that there is no space for relativism since we cannot make sense of alternative ways of "being minded", or of a culture that shares none of our certainties. I shall not here rerun Barry Stroud's (1984) familiar arguments against Bernard Williams (1971/1981). I therefore confine myself to a comment on Coliva and Pritchard. I agree with the idea that we would be unable to understand a culture that shared none of our certainties. But I do not see why this idea blocks anything but the most absurdly radical form of relativism. Relativist anthropologists, sociologists or philosophers have never denied that there may be, or indeed are, some contingent universals. But they have denied that these contingent universals suffice to determine which other certainties we ought to adopt. Coliva and Pritchard (as well as Boghossian 2006) assume that provided only that *some* certainties are shared, disagreements over other certainties cannot permanently persist amongst people who are willing to argue rationally. But why should this be so? We can agree that the earth exists

or that  $2+2=4$ . But why would agreement on these certainties help us negotiate our different certainties in the religious domain?

Details aside, what kind and strength of relativism is discernible in *On Certainty*? The most important lesson of the above is, I take it, that Wittgenstein is not trying to defend or develop a global or comprehensive form of epistemic relativism. Instead, he is trying to sensitise us to the variety of our responses in the variety of cases in which our certainties are, or seem to be, denied by others. Sometimes our response is dismissal. Sometimes our response is education. But there is also a space for epistemic relativism of a form in which Practice-centred Exclusiveness, Notional Confrontation and Symmetry as Non-appraisal loom large. Wittgenstein seems to think that this form of epistemic relativism is at least permissible when encountering disagreements at the borderline between current science on the one hand, and magic or religion, or fundamentally different conceptions of rationality, on the other hand.

To give this form of relativism a name, I propose calling it a “relativism of distance” – a term introduced by Bernard Williams in a different context (1981). Williams emphasizes two central elements in this type of relativism. First, the confrontation with the other culture is merely “notional”. That is to say, going over to the other side is not a real or “live” option for oneself. One cannot imagine adopting the view of the other side without making an endless number of changes to one’s system of beliefs or values. And second, one’s own “vocabulary of appraisal” seems out of place: “[...] for a reflective person the question of appraisal does not genuinely arise [...] in purely notional confrontation.” (1981, pp. 141–142) Take once more an encounter with a Dalai Lama who denies that the world has existed since long before his birth. This qualifies as a notional confrontation for me since to adopt his attitude towards the existence of the world would involve a complete reordering of what for me are now real options. It is hard for me to imagine what kind of argument could possibly incline me to accept this option. Moreover, my vocabulary of appraisal seems inappropriate in this encounter. I feel too distant to assess the Dalai Lama’s beliefs, and thus I have no real choice but to “put up” with his views.

## 5 Epistemic contingency and groundlessness

In the last section it emerged that versions of Dependence, Pluralism, Exclusiveness, Notional Confrontation, Symmetry, and Tolerance can all be found as elements of Wittgenstein’s relativism of distance. But I have not yet made reference to Contingency, Groundlessness and Underdetermination, and their relationship to the other elements of Wittgenstein’s relativism. In this section I turn to this task.

Note, first of all, that Underdetermination, Contingency and Groundlessness are central throughout Wittgenstein's writings from the 1930s onwards. Here are some striking paragraphs from *On Certainty*. §§§ 131 and 512 express Underdetermination, §§§ 94, 107, 166, and 336 Contingency and Groundlessness:

131. No, experience is not the ground of our game of judging. Nor is its outstanding success.  
 512. Isn't the question this: "What if you had to change your opinion even on these most fundamental things?" And to that the answer seems to me to be: "You don't *have* to change it. That is just what their being 'fundamental' is."  
 94. [...] my *Weltbild* [...] is the inherited background [...]  
 107. Isn't this [i. e. our commitments to certainties; MK] altogether like the way one can instruct a child to believe in a God, or that none exists, and it will accordingly be able to produce apparently telling grounds for the one or the other?  
 166. The difficulty is to realize the groundlessness of our believing.  
 336. But what men consider reasonable or unreasonable changes.

What role do these paragraphs and thus Underdetermination, Contingency and Groundlessness play in Wittgenstein's relativism? *On Certainty* does not say much about these connections but it seems plausible to suggest something like the following consideration.

The question is how we are affected by the realization that our epistemic system or practice is contingent, groundless and underdetermined: our common-sense certainties would have been different if history had been different; they are not justifiable as a whole; and they are not fixed by the way the world is. The discussion over the past few sections shows that Wittgenstein's answer to this question is not uniform. The effects of the realization of Contingency, Groundlessness and Underdetermination are different in different domains.

In the realm of science, it seems, the realization in question does not lead to relativism or skepticism. For instance, the thought that, had our upbringing or our circumstances been different, we would not now take it for granted that water boils at about 100C°, or that humans did not visit the moon before 1969, does nothing to weaken our readiness to believe both of these propositions. Nor does an encounter with another culture – past or present – a culture that seems to deny these certainties, lead us to think that these certainties are true-for-us (or justifiable-for-us) but false-for-them (or unjustifiable-for-them). Our natural instinctive response – the response we find ourselves compelled to give – is to credit ourselves with the *epistemic luck* that derives from having been exposed to the right kind of education needed for learning what it takes to appreciate the truth of modern science and its preconditions.

The effects of realizing that our religious (un-)believing is groundless, contingent and underdetermined are different. An atheist who recognizes that her

unbelief is groundless, contingent and underdetermined does not act irrationally if she draws the conclusion that the attitude of Non-Appraisal is a permissible intellectual response towards religious thought. This response will be particularly natural if, first, the atheist has every reason not to doubt the intellectual abilities of the believers, and second, if she feels that the way of seeing the world of the believers is not, or not fully, accessible to her. This is Wittgenstein's response in *Lectures on Religious Belief* and it also looms large in *On Certainty*.

It thus seems plausible to say that recognizing the force of *Contingency*, *Groundlessness* and *Underdetermination*, regarding our epistemic system or practice, leads Wittgenstein to a kind of relativism with respect to religion and magic. But Wittgenstein is not tempted by a relativism that extends to natural science and competing ways of explaining natural events. With respect to these areas he accepts at most a form of methodological or heuristic relativism: not to appraise, and to treat as equally correct, might be the best way of identifying what is specific about such practices.

This is still not quite the full story, however. Over the past three paragraphs I have explained that for Wittgenstein we somehow *cannot help* taking an anti-relativistic stance in the case of natural science, but that we are able to choose the option of non-appraisal in the case of a confrontation with religion or magic. The idea is that ultimately our “nature” – social and natural – will not allow for relativism when it comes to questions to do with empirical investigation of natural phenomena. Putting the point in this way ought to sound familiar. It recalls Sir Peter Strawson's one-time explanation of Wittgenstein's allegedly Humean argument against epistemic skepticism:

According to Hume the naturalist, skeptical doubts are not to be met by argument. They are simply to be neglected [...]. They are to be neglected because they are *idle*; powerless against the force of nature, of our naturally implanted disposition to belief. (1985, p. 13)

[...] In spite of the greater complication of Wittgenstein's position, we can, I think, at least as far as the general skeptical questions are concerned, discern a profound community between him and Hume. (1985, p. 19)

I am not here concerned with the question whether Strawson's interpretation of Wittgenstein's anti-skeptical argument is correct. I am rather interested in the criticism that has sometimes been levelled against this proposal. I mean the criticism according to which Strawson's idea in fact “concedes the sceptic's theoretical invulnerability” (M. Williams 1988, p. 416; cf. M. Williams 1986; Sosa 1998). By insisting that the sceptic can be defeated only on the practical ground of our de-facto inability to live skepticism, Strawson concedes that the sceptic wins the theoretical argument. As far as rational arguments go, skepticism stands undefeated.

The issue for us is not skepticism, but relativism. But the parallel should be clear. There is nothing but our (social) nature that stands between us and the sceptic (says Strawson). There is nothing but our (social) nature that stands between us and a relativistic substantive attitude towards natural science and its competitors. There is no non-circular rational argument in favor of our attitude; there is only “something animal” (§ 359), “instinct” (§ 475), “the favor of Nature” (§ 505), and “our life” (§ 559). Relativism is not defeated by a theoretical-rational argument.

## Acknowledgement

Research for this paper was made possible by ERC Advanced Grant 339382, “The Emergence of Relativism”.

## Bibliography

- Anscombe, G. E. M. (1976): “The Question of Linguistic Idealism”. In: *Acta Philosophica Fennica* 28, pp. 188–215.
- Barnes, Barry/Bloor, David (1982): “Relativism, Rationalism and the Sociology of Knowledge”. In: Martin Hollis/Steven Lukes (eds.): *Rationality and Relativism*. Oxford: Blackwell, pp. 21–47.
- Boghossian, Paul (2006): *Fear of Knowledge: Against Relativism and Constructivism*. Oxford: Clarendon Press.
- Coliva, Annalisa (2010a): *Moore and Wittgenstein: Scepticism, Certainty, and Common Sense*. Palgrave Macmillan: Houndsmills, Basingstoke.
- Coliva, Annalisa (2010b): “Was Wittgenstein an Epistemic Relativist?”. In: *Philosophical Investigations* 33, pp. 1–23.
- Grayling, Anthony (2001): “Wittgenstein on Scepticism and Certainty”. In: Hans-Johann Glock (ed.): *Wittgenstein: A Critical Reader*. Oxford: Blackwell, pp. 305–21.
- Haller, Rudolf (1995): “Was Wittgenstein a Relativist?”. In: Rosaria Egidi (ed.): *Wittgenstein: Mind and Language*. Dordrecht: Kluwer, pp. 223–31.
- Harman, Gilbert/Thomson, Judith Jarvis (1996): *Moral Relativism and Moral Objectivity*. Oxford: Blackwell.
- Kusch, Martin (2011): “Disagreement and Picture in Wittgenstein’s ‘Lectures on Religious Belief’”. In: Richard Heinrich/Elisabeth Nemeth/Wolfram Pichler/David Wagner (eds.): *Image and Imaging in Philosophy, Science and the Arts, Volume 1*, (Publications of the Austrian Ludwig Wittgenstein-Society. New Series, Vol. 16). Frankfurt am Main: Ontos, pp. 35–58.
- Pritchard, Duncan (2009): “Defusing Epistemic Relativism”. In: *Synthese* 166, pp. 297–412.
- Rosen, Gideon (2001): “Nominalism, Naturalism, Epistemic Relativism”. In: *Philosophical Perspectives* 15, pp. 60–91.
- Schmitt, Frederick. F. (2007): “Introduction: Epistemic Relativism”. In: *Episteme* 4, pp. 1–9.
- Sosa, Ernest (1998): “P. F. Strawson’s Epistemological Naturalism”. In: Lewis E. Hahn (ed.): *The Philosophy of P. F. Strawson*. Chicago and Lasalle: Open Court, pp. 361–69.

- Strawson, Peter F. (1985): *Skepticism and Naturalism: Some Varieties: The Woodbridge Lectures 1983*. London: Methuen.
- Stroud, Barry (1984): "The Allure of Idealism". In: *Proceedings of the Aristotelian Society* 58, pp. 219–242.
- Williams, Bernard (1981): *Moral Luck*. Cambridge: Cambridge University Press.
- Williams, Michael (1986): "Review of P. F. Strawson, *Skepticism and Naturalism*". In: *Philosophical Quarterly* 36, pp. 444–448.
- Williams, Michael (1988): "Epistemological Realism and the Basis of Scepticism". In: *Mind* 97, pp. 415–439.
- Williams, Michael (2007): "Why (Wittgensteinian) Contextualism is not Relativism". In: *Episteme* 4, pp. 93–114.
- Wittgenstein, Ludwig (1966): "Lectures on Religious Belief". In: Cyril Barrett (ed.): *Wittgenstein: Lectures and Conversations on Aesthetics, Psychology and Religious Belief*. Berkeley and Los Angeles: University of California Press, pp. 53–72.
- Wittgenstein, Ludwig (1969): *On Certainty*. Oxford: Blackwell. (=OC)
- Wittgenstein, Ludwig (1976): *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge 1939*. Ed. C. Diamond. Chicago: Chicago University Press.
- Wright, Crispin (2004): "Wittgensteinian Certainties". In: Denis McManus (ed.): *Wittgenstein and Scepticism*. London: Routledge, pp. 22–55.

Christian Helmut Wenzel

## Wittgenstein and Free Will

**Abstract:** In this essay I wish to do three things. First, I wish to discuss a statement from the *Tractatus* which says that our free will consists in our ignorance of future actions: “The freedom of the will consists in the impossibility of knowing actions that still lie in the future. We could know them only if causality were an inner necessity like that of logical inference.” (5.1362)<sup>1</sup> I think this statement might well be inspired by a claim Moore made in connection with free will in his 1912 book *Ethics*: “We can hardly ever know for certain beforehand, which choice we actually shall make”.<sup>2</sup> But I think Moore’s claim in favor of free will is not convincing. Second, I wish to discuss a question raised in *Philosophical Investigations*. Wittgenstein asks what remains if we “subtract” the fact that my arm goes up from the fact that I raise my arm, and he adds in brackets: “Are the kinaesthetic sensations my willing?” (621)<sup>3</sup> This added reflection is a reaction to ideas put forward by William James. Wittgenstein opposes these ideas. He argues that we should not think of the will as a cause at all, kinaesthetic or not, but rather as something embedded in, and constituted by, certain contexts of learning, expectation, practice, and lack of surprise. This is a strong claim. He also returns to the question about the predictability of the future: “When people talk about the possibility of foreknowledge of the future they always forget the fact of the prediction of one’s own voluntary movements” (629).<sup>4</sup> The question is whether Wittgenstein has solved, or dissolved, the problem of free will. Some think that this is the case. I doubt it is. This is the third point I wish to discuss.

---

1 “Die Willensfreiheit besteht darin, dass zukünftige Handlungen jetzt nicht gewußt werden können. Nur dann könnten wir sie wissen, wenn die Kausalität eine innere Notwendigkeit wäre.”

2 G.E. Moore, *Ethics* 1912, p. 219.

3 “Was ist das, was übrigbleibt, wenn ich von der Tatsache, dass ich meinen Arm hebe, die abziehe, dass mein Arm sich hebt? ((Sind nun die kinaesthetischen Empfindungen mein Wollen?))” That Wittgenstein here opposes ideas by William James about the will and kinaesthetic feelings, has been argued by Peter Hacker (1996), Stewart Candlish (2001), and John Hyman (2011). Hacker discusses James’ “ideo-motor theory” and Wundt’s “innervations theory” in this context (Hacker 2000, pp. 192–238, especially 199–208, 218–220).

4 “Wenn Leute über die Möglichkeit eines Vorherwissens der Zukunft reden, vergessen sie immer die Tatsache des Vorhersagens der willkürlichen Bewegungen.”



## 1 Free will in the *Tractatus*

In the *Tractatus* Wittgenstein is still somewhat under the spell of Schopenhauer, especially regarding the notion of “will”. Schopenhauer in turn is very much influenced by Kant. He thought that Kant’s “solution” of the problem of free will by means of the distinction between phenomena and noumena – things as they appear and as they are in themselves, was one of the greatest achievements in philosophy. Part of this “solution” is the idea of an “intelligible” or “noumenal” choice (*intelligible Wahl*), by means of which you choose the whole world (of appearances). You choose your empirical world. Through your “intelligible character” you choose your “empirical character” – and then are stuck with it. One can also say that through your intelligible choice you choose, once and for all, your empirical choices, all the moral choices and moral decisions you are going to make in the course of your life. This Kantian idea is, I think, reflected in the *Tractatus*. “If the good or bad exercise of the will does alter the world, it can alter only the limits of the world... In short the effect must be that it becomes an altogether different world” (6.43). World and will are “equivalent” to each other (*Tagebücher* 4.11.16). Another will means another world. Here it appears as if the will stands outside the world, comparable to a noumenal self. Also the solipsism that we find in the *Tractatus* can be argued to have its roots via Schopenhauer in Kant, because it is through my (solitary) “intelligible choice” that I choose my whole empirical world. This view can be found in Kant’s *First Critique*. But also in Kant’s moral philosophy, the idea of a reflection about the universalizability of maxims and the question whether I can really *will* such maxims, we find a solipsistic flavor: “Act only in accordance with that maxim through which you can at the same time will that it become a universal law”.<sup>5</sup> Even though you think of others universally, you have to think alone, by yourself. The “real” subject is not part of the phenomenal world, as the intelligible character is not part of the empirical character. Rather, the former is the noumenal “cause” of the latter.<sup>6</sup> This idea is an integral part of Kant’s transcendental theory of “causality

---

5 “Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde.” (Immanuel Kant, *Academy Edition* vol. 4, page 421) The expression “wollen kannst” (can will) is often overlooked. But I think it is essential. It makes Kant’s moral philosophy less formal and more personal. But it also gives it a solipsistic flavor. You have to think of others, but you think by yourself. You have to ask yourself whether you really *want* that maxim you are now considering. Besides universality (*can will*) there is an element of willing (*can will*). You have to ask yourself whether you really want the maxim in question.

6 Kant is here stretching the concept of causality, because strictly speaking it applies only to phenomena, and noumenal causality is not a phenomenon. Noumenal causality cannot be ex-

of freedom”, and this theory might well be, via Schopenhauer, part of the background of Wittgenstein’s statements such as “I am my world” (5.63) and “The subject does not belong to the world; rather, it is a limit of the world” (5.632). A transcendental subject that chooses the (his or her) phenomenal world can be understood as the “limit” of this world, or as “being” this world. But all this is rather speculative philosophy, and I do not want to go too deep into this kind of philosophy here.

When Wittgenstein talks of free will, which he explicitly does only once in the *Tractatus*, I do not believe he is thinking of Kant’s “causality of freedom”. In Kant this is essentially moral freedom, and it is a form of causality. But in Wittgenstein I believe there is no intended connection with morality when he says “The freedom of the will consists in the impossibility of knowing actions that still lie in the future” (5.1362). He distances himself from any kind of causality. Rather, there might be a reference to G.E. Moore, whose book *Ethics* of 1912 he might well have known, or at least he might have been familiar with the ideas expressed in it. The context of 5.1362 is an elaborate theory about propositions, truth functions, and elementary propositions (5), and the way the truth of a proposition follows from another proposition (5.13). I will briefly lay out this context. Wittgenstein claims that there is no way to draw conclusions from facts or situations [*Sachlagen*] to totally different facts: “There is no possible way of making an inference from the existence of one situation to the existence of another, entirely different [*gänzlich verschiedenen*] situation” (5.135). The question is of course how “entirely different” given situations actually are, and this becomes crucial in the sentences that follow. He immediately claims that such inferences have nothing to do with causality: “There is no causal nexus to justify such an inference” (5.136). Like Hume, he thinks that causal relations fall short of logical relations.<sup>7</sup> But the question is what “such an inference” refers to and what it does not refer to, i. e. when two situations are “entirely different”. After the introduction of the idea of a causal nexus follow immediately the statements about

---

perceived. At best the effect is a phenomenon (the world chosen), but the cause (the noumenal subject) is not. The cause is the transcendental subject as a thing in itself. It is not part of the phenomenal world. In general, for Kant things in themselves are not additional things. They are the usual phenomenal things considered as independently of time and space. In this way we should try to make sense of “noumenal causality”. The same applies to Kant saying that we are “affected” by things in themselves.

<sup>7</sup> Cyril Barrett (1996) has pointed out that Wittgenstein goes even further than Hume. Whereas Hume argued that we cannot perceive a causal nexus (but only constant conjunction), Wittgenstein, like the “Ockhamites, Nicholas of Autrecourt and John of Miracourt”, says that “from one event another cannot be deduced: there is no entailment, no necessary connection between them.” Wittgenstein goes “far beyond” the “Humean line” (p. 33).

the future and about freedom of the will with which I am concerned in this essay: “We cannot infer the events of the future from those of the present. Belief in the causal nexus is superstition [*Aberglaube*]” (5.1361), and then: “The freedom of the will consists in the impossibility of knowing actions that still lie in the future” (5.1362). There is no causal nexus, hence we cannot know the future, and hence we are free.

Does Wittgenstein subscribe to the last statement (5.1362)? Does he believe in such freedom of the will? Or does he think such belief is nonsense? This is not made explicit in that passage.<sup>8</sup> But there is a similar epistemic move in Moore’s defense of freedom of the will, and Moore does believe in such freedom of the will.

Moore in his *Ethics* of 1912 discusses free will in a separate chapter. In the book as a whole he develops a theory of right and wrong that depends on what an agent *can* do, and all that is necessary for this theory, Moore says, is an understanding of the word “can” expressing what an agent *can* do *if* he so chooses. No sense of any agent that “absolutely can” is required (198). But Moore also expresses doubt and reservation about his theory. He is not sure whether it adequately grasps what is right and wrong (202). In any case he thinks that free will is ordinarily understood to require that we have “the power of acting differently from the way in which we actually do act” (203). He thinks it is clear that we often “can, *in some sense*, do what we don’t do” (204). To show this he gives the following example: “I could have walked a mile in twenty minutes this morning, but I certainly could not have run two miles in five minutes. I did not, in fact, do either of these two things; but it is pure nonsense to say that the mere fact that I did not, does away with the distinction between them, which I express by saying that the one was within my powers, whereas the other was not” (206). To make this distinction possible, we need a “generic” meaning of “could”. Moore on other occasions has walked a mile in twenty minutes, and that justifies his saying that he could have done it this morning, too. In another sense of “could”, a “particular” sense of “could”, it might be true to say that he could *not* have done it this morning. If one is a determinist, thinking that everything has a preceding cause from which it necessarily follows, then one might think so. The word “could” is simply ambiguous. We can thus rescue free will by saying that we sometimes “could” (in the generic sense) have done what we did not do; and that, if deter-

---

<sup>8</sup> Georg Henrik von Wright (1974) said in his 1972 Woodbridge Lectures: “The idea that human freedom is relative to human ignorance is familiar from the history of thought. [...] It was given an aphoristic formulation in modern times by Wittgenstein when he wrote in the *Tractatus* (5.1362): ‘The freedom of the will consists in the fact that future actions cannot be known now.’” (p. 133) He says that Wittgenstein gave it a “formulation”. I do not know whether he thinks Wittgenstein subscribed to this formulation, or even believed in the existence of free will.

minism is true, we “could not” (in the particular sense) have done what we did not do (211). In this way, free will and determinism are compatible.<sup>9</sup>

There are two things I wish to add here. One is that G.E. Moore hints that he is ready to give up determinism and the principle that everything has a cause (pp. 209–210). The other is that he is content with the understanding of “free will” as merely requiring that we could have done otherwise if we had chosen otherwise (and that no “absolutely could” is necessary). He is content with saying that in that sense of the term we “really have” free will (p. 217). Moore thus has been struggling with the issue. But he is not yet finished. He next faces the objection that some might require that we could not only have acted differently but also have *chosen* differently. Could we have chosen what we did not choose? To this he gives two answers. One is that indeed we “should have chosen to do a particular thing if we had chosen to make the choice” and he gives some kind of evidence for this. It can for instance be seen when we make an “effort to induce ourselves to choose a particular course” (p. 219). The other answer is that “we can hardly ever know for certain beforehand, which choice we actually shall make” (p. 219). This is the sentence I wish to point out in relation to Wittgenstein’s statement in the *Tractatus*: “Freedom of the will consists in the impossibility of knowing actions that still lie in the future” (5.1362).

Moore thinks the word “possible” is “not ambiguous” (p. 209) and that this implies that “some things, which did not happen, could have happened” (p. 209). In his eyes this contradicts the principle that everything has a cause (a preceding cause that is an event), and he is ready to “give up this principle” (p. 210). This intuition underlies also his defense of our ability to freely choose among alternatives:

One of the commonest senses of the word ‘possible’ is that in which we call an event ‘possible’ when no man can know for certain that it will not happen. It follows that almost, if not quite always, when we make a choice, after considering alternatives, it was possible that we should have chosen one of these alternatives, which we did not actually choose; and often, of course, it was not only possible, but highly probable, that we should have done so.

I find it difficult to follow Moore’s intuitions at this point. He makes the notion of possibility depend on our epistemic limitations. These limitations justify our way of using the word “possible”. If nobody can know for certain that an event will not happen, we may call the event “possible”. This makes good sense, it seems to

---

<sup>9</sup> Smart (1961) makes a similar distinction and speaks of a “specific” and an “ordinary”, everyday sense of the word “can”.

me. But how do we know that nobody can know for certain that an event will not happen? This is something that worries a determinist. And why does it “follow” that it was possible that “we should have chosen one of these alternatives, which we did not actually choose”? I think he thinks of this case in a parallel way as he thinks of the generic meaning of “could”. It reflects the way we think and use words. But a determinist will not be impressed by this.

If we assume that Wittgenstein was familiar with the arguments Moore had put forward in his 1912 book *Ethics*, what conclusions can we draw regarding 5.1362, where Wittgenstein says: “The freedom of the will consists in the impossibility of knowing actions that still lie in the future”? Wittgenstein flatly dismisses belief in the principle of causality: “We cannot infer the events of the future from those of the present. Belief in the causal nexus is superstition.” (5.1361)<sup>10</sup> Moore is ready to dismiss belief in the principle of causality. But he does not flatly reject it. Wittgenstein’s criticism is much stronger and more radical than is Moore’s. It seems he subscribes to Moore’s intuition that “we can hardly ever know for certain beforehand, which choice we actually shall make” and that “no man can know for certain what will happen” (p. 219). What 5.1361 says about free will at least expresses such an intuition. But for Moore this does not define free will. Far from doing so, “it goes no way to justify us in saying that we have Free Will” (p. 219). Moore is instead relying on our everyday use of the expression “could” in the generic sense. He thinks that this usage makes sense and cannot be given up. The epistemic move about our ignorance is thus not central to his justification of free will. But for Wittgenstein the freedom of the will “consists” in this ignorance, i.e. “the impossibility of knowing actions that still lie in the future”. At least this is what 5.1362 says. Whether this implies that Wittgenstein himself believes in free will and that it “consists” in this ignorance is another question.

In the *Tractatus* we thus find on the one hand a radical dismissal of causality and on the other hand that the will and the self still play roles in ways inspired by Kant and Schopenhauer that are very speculative. Wittgenstein later distances himself from such speculative philosophy, but he keeps the dismissal of any understanding of will as having causal powers. Not much is left for will to do in the *Tractatus*. As a phenomenon it is “of interest only to psychology”

---

**10** I am not saying that Wittgenstein is influenced by Moore regarding his view of causality. The source might well be David Hume, who argued that we observe a “uniform and regular conjunction” between events in nature and that our idea of a necessity in this conjunction is merely the result of a “determination of the mind to pass from one object to its usual attendant, and infer the existence of one from that of the other” (*Treatise*, Book II, Part III, Section I *Of Liberty and Necessity*, p. 448). Thus the inference is psychological, not logical. I thank Peter Hacker for pointing this parallelism between Hume and Wittgenstein out to me.

(6.423). This dismissal is again similar to Kant, to whom rational psychology does not have much to say, and who in the first *Critique* is not interested in empirical psychology. But regarding ethics, according to the *Tractatus* it is “impossible to speak about the will” (6.423), and this is different from Kant, who wrote a whole ethical theory in which the will plays an important role. Kant talks about the good will and causality of will. But in the end, even Kant has little to say about the will understood as cause, because as free will it must be noumenal and not much can be said about the noumenal self. Kant insists on the possibility of “causality of freedom”, but this remains speculative and empty in spite of all his efforts to give it a more substantial and positive meaning. The third part of the *Groundwork* clearly shows how Kant cannot fully achieve what he set out to do (prove the existence of free will). Wittgenstein in 5.631 writes “There is no such thing as the subject that thinks or entertains ideas”. Kant in the *First Critique* develops a transcendental account of “original synthetic unity of apperception” and in his Lectures a few years before the publication of the *First Critique* he is reported to have said that “When I say: I think, I act, etc, then either the world “I” is used wrongly or I am free”.<sup>11</sup> This sounds more optimistic and similar to Moore and his analysis of how we use the word “could”. But this 1778–9 argument is not taken up in the 1781 *First Critique*. In general, Kant became increasingly skeptical about how much we can justifiably say about freedom of thought, freedom of will, and freedom of action. But he did not give up his moral theory and what he says about good will.

Contrary to Moore and possibly Wittgenstein, I think the epistemic defense of free will that is based on our ignorance of the future is not a strong defense. Spinoza had already attacked this view severely in the 17<sup>th</sup> century.<sup>12</sup> It is true

---

11 “Wenn ich sage: ich denke, ich handle, usw.; dann ist entweder das Wort Ich falsch gebraucht, oder ich bin frei ... Ich thue, als actio, kann nicht anders als absolute frei gebraucht werden. Alle praktischen Sätze hätten keinen Sinn, wenn der Mensch nicht frei ware.” (Metaphysics Lectures Pölitz, around 1778–9, *Academy Edition* vol. 28, p. 269)

12 Cyril Barrett (1996) observes that “Spinoza and Wittgenstein are saying opposite things” (p. 36). Spinoza is a determinist believing in causality and thinking that our belief in free will is an illusion. Wittgenstein on the contrary thinks it is our belief in causality that is an illusion (*ein Aberglaube*), and that, as Barrett reads Wittgenstein, “it is because we are free that neither we nor anyone else can predict what we are going to do until we decide to do it” (36–7, footnote 4). But differently from Barrett, I see an unresolved tension, if not a contradiction, between this deciding to do something (apparently in the world) on the one hand, and the idea that “the world is independent of my will” and that I am “completely powerless” on the other (p. 34). The problem comes up, I think, also in Barrett’s discussion of Anscombe’s criticism. Anscombe says that Wittgenstein seems to her to get it “obviously wrong” (p. 35) when he talks about the relationship between action and intention. Barrett defends Wittgenstein by

that if a Laplacian demon could predict our future action, he should better not tell us, because then we could do otherwise and contradict his prediction. But if such a demon keeps the knowledge to himself or if there are natural laws, unknown to us but according to which our future actions are fixed, free will seems empty. This is so even if Moore's analysis or our ways of talking of abilities and our use of the phrase "he could (in the generic sense) have done otherwise" still make sense. In the *Tractatus*, will as causal power is rejected, but in a transcendental sense it still plays a role. What status "free will" actually has, or might have, is not said.

## 2 Free will in the Philosophical Investigations

In the *Philosophical Investigations* Wittgenstein discusses will and willing in sections 611 to (around) 641. This is not a separate topic. It is embedded in continuous discussions of various mental acts and mental states such as expecting, feeling, meaning, remembering, recognizing, describing, and intending. We often lack the words for describing feelings and things, for instance the aroma of coffee (610). It is immediately after such a topic as describing the aroma of coffee that Wittgenstein begins to talk about the will. He usually speaks of the "act of willing" (*das Wollen*), but sometimes also of the "will" (*der Wille*). The dismissal of any kind of causal account of will that we found in the *Tractatus* can also be found in the *Investigations*. But instead of a Kantian and Schopenhauerian transcendental account of self and will, we find emphases on contexts, situations, and familiar practices. This I think is the basic shift in comparison with the *Tractatus*. Speculative, transcendental accounts and reflections are completely out of the picture now. Nevertheless, several examples that Wittgenstein thought about in his *Notebooks* during October and November 1916 appear again in the *Philosophical Investigations*, such as the example of lifting one's

---

saying on the one hand that the intention is part of the action (p. 37) and on the other that "the action itself" (p. 36) has no moral value and is independent of the intention. But I cannot see how he can have it both ways. It seems to me that without the intention an action is not an action any more. Wittgenstein wants to keep the will, especially the ethical will, completely out of the world. But how can Barrett then make sense of the idea of "renunciation" (p. 39) as a way of achieving freedom of the will? It seems to me that an act of renunciation is part of this world. After all, it has effects in the world, as does our thinking and desiring. Kant runs into similar problems when he tries to make sense of the freedom of the will in his moral theory.



arm, or deciding to do so in five minutes.<sup>13</sup> Wittgenstein is here conversing with his earlier self. In the *Notebooks* he wrote of *Willensakt*, *Wollen*, and *Wille*. In the *Investigations* he only writes of *Wollen* and *Wille*. There is no talk of *Willensakt* any more. The reason might be that he has abolished any idea of a separate act of will that precedes and causes voluntary actions so thoroughly that he does not even speak of *Willensakt* any more. But he still writes of voluntary acts as *willkürliche Handlung* (616, 627) and *willkürliche Bewegung* (614, 629). “Willkürlich” contains the word “Wille”, but this is problematic.

The German expression “*willkürliche Bewegung*” is translated as “voluntary movement”. This makes good sense, because in English we say “I move my arm”, and in German we say “*Ich bewege meinen Arm*”. But movement, or motion, is not an action. John Hyman has criticized Wittgenstein for confusing the two: “Wittgenstein’s second major mistake was his failure to distinguish between action and motion, or to consider how they are related” (Hyman 2011, p. 466). “[L]ike many of his predecessors, including Schopenhauer, he confuses action and motion” (467). But I do not think Wittgenstein made such a mistake. In the example Hyman is focusing on (*Investigations* 627), Wittgenstein considers two sentences: “I form the decision to pull the bell at 5 o’clock, and when it strikes 5, my arm makes this movement,” and “See, my arm is going up.” Hyman is right in saying that what is wrong with the second sentence is not only that the surprise (“and see!”) is out of place but that also that the expression “my arm is going up” is inappropriate. I raise my arm when I pull the bell. It does not “go up”. Wittgenstein points out the former (that there is no surprise), but not the latter (that we should say “I raise my arm” and not “My arm goes up”). Hyman thinks Wittgenstein was unaware of the latter. But I doubt this is the case. Why should he have been unaware of it? I see no evidence for this. Rather, I think Wittgenstein was content with pointing out the former mistake (describing the situation as if it were a surprise). He does not need to point out the latter mistake too. Thus I think it is wrong to say of Wittgenstein: “It does not occur to him that ‘my arm is going up’ describes the motion of my arm, and that the action and the motion – the *Handlung* and the *Bewegung* – are different things” (Hyman 2011, p. 467). To the contrary, I believe he knows very well that these are two different things. Although it is true that Wittgenstein “shifts back and forth between *Handlung* and *Bewegung*” (467), I do not think he does so “without considering whether there is a difference between the two” (ibid). Firstly, I think he is aware of the difference and very much interested in

---

<sup>13</sup> See *Notebook* entries 20.10.1916 and 4.11.1916 and *Philosophical Investigations* 612, 614, 624, 625, and 627.



it, because he wants to bring out what we mean by talking of the voluntariness of an action as if it were some kind of mystical inner event or action. Secondly, language is not reliable here. We say “*Ich bewege meinen Arm*”, although this is an action and not just a movement, or motion. So he is justified in sometimes using the word “*Bewegung*”.

In 621, Wittgenstein asks “What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm?” The difference is not some kind of kinaesthetic sensation, as William James has suggested. It is not my trying either. Nor is it my deciding. The difference is – this is Wittgenstein’s new suggestion – one of context, learning, expectation, familiarity, lack of surprise. John Hyman has illustrated this well: “For example, we regard the movements of a child playing with a doll as voluntary, not because we postulate invisible acts of will or images of kinaesthetic feelings preceding them and making them occur, but because we know that the child has learned how to make these movements, because the movements are coordinated and purposeful, because the child attends to what it is doing, is not alarmed or surprised or distressed by its own movements, and so on. If we feel the need to postulate a hidden cause, it is because we ignore these features of a movement and its context, which are for the most part in plain view” (459). Hyman also argues that there is voluntary passivity and not just voluntary activity and that this has usually been overlooked. “Children are sometimes picked up and carried voluntarily” (460). They let it happen and do so voluntarily. I think we may say that letting it happen is an action of some sort, too. But in any case, all this fits well into what Wittgenstein is after, namely pointing out contexts instead of hidden causes.

Wittgenstein focuses on examples like raising one’s own arm, or getting out of bed (*Brown Book*, p. 151). He questions our understanding of will and willing as causal events or causal powers, and he questions, or even dissolves such an understanding by pointing out the contexts in which the actions occur. He argues that such contexts are all that is required to understand what will and willing are. But we should notice that when he discusses voluntary actions, he mostly talks of moving arms and limbs. He does not talk of moral actions. As he did in the *Tractatus*, he stays away from ethics. This is very different from Kant, who thinks we have free will mainly in moral actions and who is not interested in any other kind of (empirical or psychological) will. This is also different in Aristotle, who begins book III of his *Nicomachean Ethics* with a discussion of virtue by observing that when we praise and blame we presuppose voluntariness:

Since virtue is concerned with passions and actions, and on voluntary passions and actions praise and blame are bestowed, on those that are involuntary pardon, and sometimes also pity, to distinguish the voluntary and the involuntary is presumably necessary for those

who are studying the nature of virtue, and useful for legislators with a view to the assigning both of honours and punishment. [Translation Sir David Ross.]

We see that in Aristotle voluntary passions are included, not just actions.<sup>14</sup> Furthermore, voluntariness is related to blame and praise, whereas Wittgenstein merely focuses on moving arms and legs without any ethical contexts. The Greek expression Aristotle uses is rendered into English as “voluntary” and into German as “*freiwillig*” (or *wollend, absichtlich*). It is not just will but *free* will that is at issue here. Free will is required for praise and blame, honors and punishment, not just will. It seems to me this context is lost in Wittgenstein.

Aristotle goes on by saying:

Those things, then, are thought involuntary, which take place under compulsion or owing to ignorance; and that is compulsory of which the moving principle is outside, being a principle in which nothing is contributed by the person who is acting or is feeling the passion, e.g. if he were to be carried somewhere by a wind, or by men who had him in their power.

The context is very wide here from the start. But at the same time there is talk of “moving principles” and of such principles being “outside” and “inside,” depending on how much the “person” is contributing. Thought about moving principles historically led to discussions of causality and determinism, which Wittgenstein wants to talk us out of. But the context of person and ethics he leaves out.

Wittgenstein wants to do away with any kind of causal explanations of the will, or free will, and he is often skeptical about talk of the inner. The discussion regarding voluntariness gets thin, because he excludes ethical contexts. Aristotle did none of this. He wrote about blame and praise. Kant preceded Wittgenstein in the former (finding problems with free will and the inner), but he did not precede Wittgenstein in the latter (excluding morality). Kant wrote a moral philosophy. Will, considered in the *Investigations* without the context of ethics, threatens to shrink to a mere point: “*Doing* itself seems not to have any volume of experience. It seems like an extensionless point, the point of a needle” (620). At least this is true when we search for a momentary act of will as first cause and phenomenon. We easily become uncertain about ourselves when we question our intentions, for instance what we wanted to say or do at a certain moment in the past. “How can you be certain that for the space of a moment [*einen Augenblick lang*] you were going to deceive

---

<sup>14</sup> “Passions” is only one of several possible translations of πάθη. Others are “emotions”, “feelings”, and “affections”. I thank John Hyman for pointing this out to me. πάθη usually indicates a negative feeling or state of mind, something unfortunate that happens to you. πάθη is different from action (πράξις).

him? Weren't your actions and thoughts much too rudimentary?" (638) Thus Wittgenstein keeps emphasizing the necessity of epistemic contexts and concrete situations to understand what we think of as inner experiences such as wishing, willing, and intending. "Then the 'inner experience' of intending [*des Wollens*] seems to vanish again. Instead one remembers thoughts, feelings, movements, and also connexions with earlier situations" (645). The situations considered are epistemic and not moral. Will and free will are here not related to what we mean by being a "person". When Wittgenstein wants to correct our "wanting to think of willing as an immediate non-causal bringing-about" (613), he is still struggling with his own former self when he was under the influence of Schopenhauer and Kant, for whom the phenomenal world was causally determined and "causality of freedom" somehow had to make sense and "bring about" something in this world. But since Wittgenstein does not believe in this kind of transcendental philosophy any more, willing and voluntariness completely disappear as acts: "'Willing' is not the name of an action; and so not the name of any voluntary action either" (613). Instead, it must be part of the action that is carried out and not an action that precedes it and takes place in the mind, as John Locke thought of it.

It is in this context that we must also understand what Wittgenstein says about the predictability of the future. In the *Tractatus* he said: "The freedom of the will consists in the impossibility of knowing actions that still lie in the future" (5.1362) (and I am not sure whether this was really his own opinion then or not), and this epistemic understanding of free will was put forward in the context of a general opposition to any kind of belief in causal connections as necessary. Now in the *Investigations*, in the context of lifting one's arm and pulling a bell (627), he notes: "When people talk about the possibility of foreknowledge of the future they always forget the fact of the prediction of one's own voluntary movements" (629). This sounds as if there were a special first-person access and certainty available in this case. Maybe I can predict what I will do next with more certainty than I can predict a solar eclipse? But I think this is not what Wittgenstein wants to say here. We can see this if we consider the context in which 629 stands. "*Doing* itself [...] seems like an extensionless point, the point of a needle" (620), and willing is not a "kinaesthetic sensation" (621). He questions such momentary certainty. A moment of time (*ein Augenblick*, 636, 638, 642) is as extensionless as the "point of a needle" (620)! We need more context, temporal as well as spatial context. Thus he inserts the observation that when we touch an object with a stick, we have "the sensation of touching in the tip of the stick, not in the hand that holds it" (626), which is an observation about our feeling regarding space; and he discusses recognition (625), deciding to do something at 5 o'clock (627), prediction (630–632), and memory (633–645), which are all observations involving time. To understand willing we

need to broaden the view in both time and space. The theme is always that we need contexts, antecedents, and stories (*eine Vorgeschichte* 638). If you recall what you felt or wanted to do, you need to recall the circumstances. Thus also “the prediction of one’s own voluntary movements” (629) depends on knowledge of familiar circumstances. One is not surprised that one raises one’s arm and the bell sounds at 5 o’clock, because one has done similar things on similar occasions. There is no act of will that causes the movement of the arm. One does it voluntarily, because one does not have a divided mind about it. One is familiar with the action, one has learned how to do such a thing, and one does not regret. All this brings us back to Aristotle again.

Thus if we combine the observation about free will in the *Tractatus*, that “freedom of the will consists in the impossibility of knowing actions that still lie in the future” (5.1362) with the discussion of will and predictability (but not “free” will) in the *Investigations*, we arrive at the suggestion that if there is such a thing as free will, it must be understood in the wider contexts of practices in our daily lives, in the various contexts of promising, predicting, expecting, recalling, intending, and expressing what we want.

### 3 Concluding remarks

Wittgenstein does not much deal with free will. But he is concerned with will. This is peculiar, because free will plays a central role in Kant and Schopenhauer who had a large influence on Wittgenstein. Why is that? I think the main reason is that Wittgenstein from the start excluded ethics, at least in his writings. He preferred to keep quiet about it and tended to equate it with aesthetics and religion. Such conflation and *Horizontverschmelzungen* can lead to blind spots. He wrote about the soul and delicate problems such as meaning-blindness. But did he have a blind spot for ethics? If one prefers to be quiet about something, there is always the danger that one lives with one-sided pictures and it will not be easy for others (like us) to judge what such a person (like Wittgenstein) really thinks.

In Aristotle the problem of free will and voluntariness (ἑκούσιος is translated as “voluntary” and “*freiwillig*”) was tied to virtue, blame, and praise. For that it does not matter whether will is understood as a causal power or not. Aristotle later talks about the source or “moving principle” of an action, and he distinguishes the case where it is in the agent from the case where it is outside of the agent. But this comes later. First we observe that people praise and blame each other for what they do. Depending on whether the act was done voluntarily or not (or the passion or emotion had voluntarily or not), we excuse, pardon, pity, or we don’t. This aspect, it seems to me, was not in Wittgenstein’s focus

of interest from the start, and this has consequences regarding to what extent we can say he was dealing with the problem of free will at all, even though he was thinking about the problem of what will and willing is.

According to Aristotle we are acting involuntarily, or at least non-voluntarily, if we act under compulsion or with ignorance. If a tyrant threatens you and orders you to do something, you might be said to do it involuntarily and you might be pardoned or pitied. If you are in a storm on a ship, you might involuntarily throw goods overboard hoping that this will make it less likely that the ship will sink. Many actions are “mixed” as Aristotle says. Partly, or in certain respects, we do them voluntarily, partly we do them involuntarily. It also matters whether we regret what we have done or not. It matters how much control we have over the action. In so far as “the moving principle” is in the person and in his power, the action or passion, feeling, or affection is voluntary. In so far as it is outside of the person, it is involuntary. Such considerations can be very complicated in particular cases and will involve us in questions of character, person, personal identity, ascription, justice, human rights, and many other topics. If we see free will in these respects, it becomes more understandable that Kant thought we can be free only in the light of morality. But Kant was also under the influence of Newton and the natural sciences. Regarding the phenomenal world, he understood the “moving principle” in the person as a form of natural causality, which he thought to be deterministic, where every cause has another cause. Hence he tried to save freedom by distinguishing phenomena and noumena. Schopenhauer followed him, and Wittgenstein tried to get out of this. But he never held on to the ethical aspect, which was the starting point in Aristotle and still strong in Kant and Schopenhauer, be it in form the categorical imperative or in the form of compassion. Thus by exorcizing the causal understanding of the will, free will went overboard as well. Wittgenstein saw free will only as will, which was a one-sided picture. Wittgenstein in the *Investigations* rightly points out contexts and circumstances, but limits himself to epistemological questions about knowledge and certainly and does not consider moral questions. He talks about drill and *Abrichtung*, not moral education. In the light of this observation we can see another source of Hyman’s criticism. I do not think that Wittgenstein does not see the difference between action (*Handlung*) and movement (*Bewegung*). He does distinguish between the two. He sees the epistemic differences. But I think he abstracts from the moral dimension and thereby misses the moral differences which might be the most important ones.

One might object to my criticism of Wittgenstein by saying that we can study questions about will and free will independently of questions about morality. One might say that such questions come first and morality can be discussed

later.<sup>15</sup> This may be true. But it might also be false. I feel uncomfortable with excluding something that was so important to Aristotle and Kant for instance. Also Peter Strawson in his discussion of freedom and resentment focuses on our reactions in moral contexts. Without the ethical dimension something is lost, and it seems to me that what is lost might be essential.

For comments and corrections I wish to thank Thomas Bernöster, Peter Hacker, Herbert Hanreich, John Hyman, Kai Marchal, Stephen Priest, and Joel Schickel.

## Bibliography

- Barrett, Cyril (1996): "Wittgenstein on Freedom of the Will". In: *Wittgenstein and Philosophy of Culture. Proceedings of the 18<sup>th</sup> International Wittgenstein Symposium, 13<sup>th</sup> to 20<sup>th</sup> August Kirchberg am Wechsel 1995, Austria*, Vienna: Hölder-Pichler-Tempsky, pp. 33–40.
- Candlish, Stewart (2001): "The Will". In: Hans-Johann Glock (ed.): *Wittgenstein: A Critical Reader*, Malden: Blackwell, pp. 156–173.
- Candlish, Stewart/Damnjanovic, Nic (2013): "Reasons, Actions, and The Will: The Fall and the Rise of Causalism". In: Michael Beaney (ed.): *Oxford Handbook of the History of Analytic Philosophy*, Oxford: Oxford University Press, pp. 689–708.
- Hacker, Peter (1996/2000): *Wittgenstein: Mind and Will. Volume 4 of an Analytical Commentary on the Philosophical Investigations*. Malden: Blackwell. As paperback in two separate volumes, *Part I Essays* and *Part II Exegesis 428–693*, Malden: Blackwell.
- Hume, David (1984): *A Treatise of Human Nature*. New York: Penguin Classics.
- Hyman, John (2011): "Action and the Will". In: Oskari Kuusela and Marie McGinn (eds.): *The Oxford Handbook of Wittgenstein*, Oxford: Oxford University Press 2011, pp. 451–471.
- Moore, G.E. (1912): *Ethics*. Oxford: Oxford University Press.
- Savigny, Eike von (ed.) (1995): *Wittgenstein über die Seele*, Frankfurt a. M.: Suhrkamp.
- Savigny, Eike von (1996): *Der Mensch als Mitmensch. Wittgenstein's Philosophische Untersuchungen*. Munich: Deutscher Taschenbuch Verlag.
- Schneider, Hans Julius (2009): "Das musste ja so kommen! Libertinische Willensfreiheit und Handlungsdetermination". In: *Erwägen, Wissen, Ethik* 20/1, pp. 57–58.
- Schröder, Severin (2006): *Wittgenstein. The Way Out of the Fly Bottle*. Cambridge: Polity Press.
- Smart, J.J.C. (1961): "Free Will, Praise and Blame". In: *Mind* 70/279, pp. 291–306.
- Wittgenstein, Ludwig (1963 [1961]): *Tractatus Logico-Philosophicus*. Translated by D.F. Pears and B.F. McGuinness, London: Routledge and Kegan Paul.
- Wittgenstein, Ludwig (1958 [1953]): *Philosophical Investigations*. translated by G.E.M. Anscombe, Malden: Blackwell.

---

<sup>15</sup> Peter Hacker and Stephen Priest pointed this out to me. They made this criticism. Indeed, Kant wrote his *Critique of Pure Reason* before he wrote his *Critique of Practical Reason*, and he discussed transcendental, cosmological freedom first independently from morality. But I think he wrote about the former with the latter in mind.

Wittgenstein, Ludwig (1960 [1958]): *The Blue and Brown Books*. New York: Harper Torchbooks.

Wright, Georg Henrik von (1974): *Causality and Determinism. Woodbridge Lectures Delivered at Columbia University in October and November of 1972*. New York: Columbia University Press.

Wright, Georg Henrik von (1985): "Of Human Freedom". In: S. M. McMurrin (ed.): *The Tanner Lectures on Human Values*, Salt Lake City: University of Utah Press, pp. 107–170.

Joachim Schulte

## Wittgenstein's Last Writings

**Abstract:** This sketch of a survey of Wittgenstein's manuscript writings between ca. 1946 and his death in 1951 attempts to do two things: (1) some comments are made on the basis for, and the possible justification of, editorial decisions taken by Wittgenstein's heirs; (2) a few topics are identified as particularly characteristic of Wittgenstein's late thought and selected for more extended philosophical discussion. Here, the chief question raised is whether, in Wittgenstein's eyes, a satisfactory distinction can be drawn between 'empirical' and 'logical' propositions.

To draw a sketch of the later Wittgenstein's philosophical attitude one will be well-advised to begin with an examination of his writings. True, there is plenty of material in addition to his manuscripts and typescripts, such as records of conversations and notes taken by students during or after his lectures as well as other kinds of reports by people who knew him. But experience will show again and again that Wittgenstein's own writings can be relied on in a completely different and evidently much more dependable way than those other types of record.

The period about which I want to talk in this paper starts with the completion of the typescript of the final version of *Philosophical Investigations*. While the exact time at which this typescript was finished is not known, we can be sure that its last page was dictated and typed sometime in the first half of the year 1946. This is the same time at which Wittgenstein was (as far as we know) busy discussing problems belonging to the philosophy of psychology. Now, this expression "philosophy of psychology", which may sound a little unfamiliar to modern ears, is Wittgenstein's own term. To be sure, most philosophers would be content to style this sphere of discussion "philosophy of mind". But Wittgenstein for his part does not use this name. An additional consideration may concern the two main sources on which Wittgenstein draws in his lectures and writings, viz., William James and Wolfgang Köhler. For, in spite of their obvious philosophical interests, these two authors saw themselves primarily as *scientific* psychologists for whom empirical methods of research were second nature.

But, however relevant this consideration may be, from 1946 to the summer of 1947 — that is, up to the time at which Wittgenstein withdrew from his official teaching post — there are many unmistakable parallels to be found between his lectures and his written work. Great parts of these lectures are extremely well documented, which renders it easy to make comparisons between Wittgenstein's writings and these lectures. But just because they are so easy to compare readers of these records will feel that what they are confronted with are not only two dif-



ferent styles of presenting the same, or closely related, material but two fundamentally different approaches. And the extent to which they differ may be such that one may find it difficult to speak of “the same, or closely related, material”. In view of this state of things it may appear natural to arrive at the following conclusion: on the one hand, Wittgenstein took his obligations as an academic teacher very seriously indeed and, in his lectures, spared no effort to get certain ideas, problems and ways of dealing with these problems across to his students. On the other hand, in his writings he was not in the least concerned to render plausible the contents of certain doctrines or views or theses. No, he declared himself a radically anti-dogmatic thinker who was chiefly concerned to show what is problematic about typically philosophical ways of thinking about various supposedly philosophical kinds of question and to cast a critical eye on claims regarding the legitimacy of such kinds of question.

Once one sees the enormous diversity between the work of an academic teacher, on the one hand, and a philosophical writer, on the other, it will not be difficult to understand that at one point or another the conflict between these two ways of proceeding was no longer bearable. It seems likely that Wittgenstein gained the impression that his work as a teacher had damaging effects on his work as a writer. And as he evidently preferred to promote work on his writings, it was surely in accord with this preference to decide that he should retire from his professorial chair and get on with his manuscripts. As a result, he moved to the solitariness of remote Irish cottages to do some writing.

At any rate, the greater part of the year 1948 as well as the first half of 1949 were spent in Ireland, partly in County Wicklow, partly in Galway, and partly in Dublin, where Wittgenstein regularly saw his friend and former pupil M. O’C. Drury. It was in Ireland that Wittgenstein completed a series of notes written down in nine consecutive manuscript volumes, apparently displaying a degree of continuity and unity of subject matter which in an author of Wittgenstein’s kind may seem almost astonishing. Most readers will know some of the central themes from a fragment on the philosophy of psychology which was first published as what used to be called “Part II” of the *Investigations*. This fragment consists of an extract from the nine volumes mentioned before — an extract Wittgenstein selected in the spring of 1949. Here, he deals with notions like “grief” or “hope”, with questions of perception, in particular the perception of aspects, as well as connected problems concerning what Wittgenstein calls meaning-blindness. And this notion raises a number of questions that can be seen to be of considerable relevance to certain issues in the field of aesthetics, for instance.

This way of emphasizing problems belonging to the province of the philosophy of psychology must not mislead us into overlooking the fact that at the same time Wittgenstein was concerned with questions of quite different kinds

such as what he called Moore's paradox, for example. What he meant by this is the obvious but not easily explained absurdity of utterances like "It's raining, but I don't believe it". From Wittgenstein's point of view, this paradox is bound up with Frege's idea of separating the assertoric force from the propositional content of a given assertion. And here it is not difficult to see that it is also bound up with the very idea of propositional attitudes, and hence with an idea which in its turn forms an essential ingredient in all kinds of speech-act theory.

Yet another topic addressed by Wittgenstein in this context is characterized by questions concerning the appropriateness or the relativity of the formation of our concepts. This is a topic closely connected with ideas entertained by Wittgenstein in the late 1930s and the first half of the 1940s in the context of his remarks on the foundations — or rather: the dispensability of foundations — of mathematics. Thus we see that, even if it is correct to say that in his late writings relatively new ideas like his reflections on seeing aspects stand at the centre of his attention, these ideas are at the same time embedded in questions and insights which — as for example his reflections on Frege's judgement stroke — reach as far back as Wittgenstein's earliest philosophical notebooks.

Many authors agree in seeing the summer of 1949 as a kind of break in Wittgenstein's life and thought. And surely this way of looking at the matter is not unjustified. For, first of all, this is the time at which the aforementioned sequence of nine volumes on the philosophy of psychology comes to an end. Second, in July Wittgenstein embarks on his second voyage to America, where on this occasion he will spend almost three months in the home of his friend and former pupil Norman Malcolm. Here, on the one hand, he experiences a dramatic decline in the state of his health while, on the other, he gets involved in discussions about the topic "certainty" and, in particular, about Malcolm's response to G. E. Moore's "Defence of Common Sense" and "Proof of an External World". Both events are, if one wishes to put it this way, signs of things to come. For, first of all, after his return from America to Cambridge Wittgenstein learns that he is suffering from an incurable form of cancer and, second, in many of his remarks written down during the last year and a half of his life he deals with questions that can be seen as deriving directly or indirectly from those papers by Moore and, in particular, from Malcolm's perspective on these papers.

Soon after his having learned about his own incurable disease in November 1949 Wittgenstein undertakes his third post-war journey to Vienna, where he arrives shortly before Christmas to spend time with his oldest sister in her last illness. He stays in Vienna for three months before returning to England in the spring of 1950. At first he lives in Cambridge for a short while; afterwards he moves to Oxford to live in the house of his former pupil Elizabeth Anscombe. Still in Vienna, he happens to come across a copy of Goethe's *Theory of Colours*,

which stimulates him to write down first notes on this topic on a few loose double sheets of paper. There is a somewhat greater number of sheets of the same type filled with notes on questions that were raised in the course of Wittgenstein's discussions with Malcolm in the American summer of 1949. This latter group of remarks is what we know as §§1–65 of the volume *On Certainty*. The former group was published as part II of the book *Remarks on Colour*.

In this way, that is, by pointing out the fact that the editors of Wittgenstein's writings tore his Viennese notes apart in the fashion described I have reached one leitmotif of my considerations. This is the question whether and, if so, to what extent this way of editing his notes has shaped our picture of Wittgenstein. But before getting down to discussing this question I shall briefly conclude the story of Wittgenstein's life.

After his return from Vienna he, in spite of his fragile state of health, continues to work with some regularity on his manuscripts. Results of this work are the so-called third part of *Remarks on Colour*, §§66–299 of *On Certainty* as well as a certain number of remarks that were published by Wittgenstein's editors in volume II of *Last Writings on the Philosophy of Psychology* (subtitled *The Inner and the Outer*) or in *Culture and Value*. Possibly, the selection of remarks printed as part I of *Remarks on Colour* was also put together during this period. Shortly before the end of the last-mentioned sequence of remarks published in *On Certainty* we encounter the first remark in this book bearing a date (§287). It was written on 23 September 1950 and speaks about squirrels that get along without relying on a law of induction. It would not surprise me if this remark had been provoked by relatives of those squirrels about which he repeatedly discoursed while walking across Christ Church Meadow in the company of Bouwsma (cf. Bouwsma 1987, pp. 66–7). A few days later he embarks on his last trip to Norway, from which he returns to England in the middle of November. This autumn journey and his subsequent state of health make it impossible for him to continue working on his manuscripts. But all of a sudden, on Saturday, 10 March 1951, he begins to write down a series of almost daily remarks which only terminates on 27 April, that is, two days before his death. Evidently, this flow of thoughts came quite as a surprise to Wittgenstein himself, who almost regarded it as a kind of miracle. For on 16 April, less than two weeks before his death, he writes in a letter to Malcolm:

About a month ago I suddenly found myself in the right frame of mind for doing philosophy. I had been *absolutely* certain that I'd never again be able to do it. It's the first time after more than 2 years that the curtain in my brain has gone up. – Of course, so far I've only worked for about 5 weeks and it may be all over by tomorrow; but it bucks me up a lot now. (16 April 1951; WiC, p. 479)

To sum up: The writings extant from the period we are looking at comprise the following: first, there is a sequence of almost nine volumes from Wittgenstein's last Cambridge terms and the year and a half spent in Ireland. In addition, there are two or three smaller notebooks whose contents were partly but by no means entirely absorbed into those nine big volumes. The second group of papers we have consists of six or seven middle-sized manuscript volumes from the time between Wittgenstein's return from America and his death. The first point to make is that neither of these two sequences of writings has, even approximately, been published in the form in which they have come down to us. Of course, all this material can be found and consulted in the Bergen Electronic Edition of Wittgenstein's *Nachlass*. But what we know of the first group in printed form are only two volumes of *Remarks on the Philosophy of Psychology*, one volume of *Last Writings on the Philosophy of Psychology* and the fragment *Philosophy of Psychology* (the former Part II of the *Investigations*). Most of these publications contain material selected by Wittgenstein himself and dictated by him to a typist.

The editors' decision to publish, not the complete sequence of nine manuscript volumes, but only the selections mentioned may have been based on the following train of thought: "The remarks selected by Wittgenstein himself must, in his eyes, have been of better quality than the remarks he discarded. Hence, printing this discarded material would go against his more or less explicit wishes. Consequently, we should publish Wittgenstein's selections." — Expressed in this abstract fashion, this argument may seem convincing. But quite apart from the fact that Wittgenstein's motives in "discarding" certain remarks may have been much more complicated than envisaged here, people familiar with the manuscript material will hasten to reply that by proceeding in this fashion an editor will deprive prospective readers of these texts of two crucial advantages:

- (1) Readers will be prevented from gaining a comprehensive overview of what Wittgenstein wrote at the time in question.
- (2) One must not forget that only familiarity with the manuscript material in its entirety will allow readers to retrace the development of Wittgenstein's thought. That is, only if you know it all will you have a chance of grasping his reasons for moving from A to B and of forming a reliable judgement about the question whether or not he actually had a good reason for the transition from A to B. In other words: if a number of links in Wittgenstein's original chain of reasoning are missing, it will in some, or many, cases be impossible to arrive at an accurate and trustworthy picture of his reasoning.

Naturally, if you go into the details of the matter, the story is much more complex and complicated than my rough exposition. But I hope that the basic thought

has become clear: without taking into account the complete text of the aforementioned nine volumes you cannot form a reliable picture of Wittgenstein's ideas — his arguments pro or anti certain views, his reconstructions and criticisms of typical ways of thinking —, to say nothing of the significance of his allusions, puns and metaphors.

A further point that should be mentioned will prove particularly relevant in dealing with Wittgenstein's writings from the last year and a half of his life. This is the fact that the volumes published have been given an orientation towards certain subject matters. By placing the volumes of the *Remarks on the Philosophy of Psychology* and the *Last Writings on the Philosophy of Psychology* through this choice of titles squarely in the ambit of the philosophy of mind readers will be confirmed in their tendency to forget about or overlook certain connections between some remarks from the last five years of Wittgenstein's life and earlier writings. Here, I only wish to point out what may seem to be a minor example. But if you think about it, it will soon gain in importance. A theme which occurs again and again in some manuscripts from the relevant time is the possibility given to man to simulate emotions and sensations and in this way to create an impression which does not correspond to reality (if you want to put it this way). One context in which Wittgenstein treats this topic is the context of his reflections on concepts and concept-formation. Thus, at one point after having discussed various aspects of our notion of pretending to be in pain Wittgenstein says:

One has to look at the concepts 'to be in pain' and 'to simulate pain' in the third and first person. Or: the infinitive covers [[is backed – or supported – by]]<sup>1</sup> all persons and tenses. Only the whole is the instrument, the concept.

But what then is the point of this complicated thing? Well, our behaviour is damned complicated, after all. (LW II, p. 37)

True, some of the things Wittgenstein says at this point are well-known from other writings of his. But our understanding of his words will not remain unaffected by the contextualisation resulting from the editors' decision to call the volume from which this quotation is taken "*Last Writings on the Philosophy of Psychology*" and to add as a subtitle "*The Inner and the Outer*".

This straightening of readers' perspective can easily distort their view. Thus, they may not only lose from sight that here there are connections with certain general as well as more specific questions. One of these general questions would be: to what extent may Wittgenstein's way of looking at things be informed by instrumental images or instrumentalist prejudices? A more specific

---

<sup>1</sup> [[...]] denotes alternative translations in this contribution.

issue would concern the highly interesting question of what one may want to call the “shape” of our concepts — and this is a question Wittgenstein mentions in a number of other manuscripts from this period. Thus, he speaks in more or less metaphorical ways of the “slopes” and “inclinations” of our concepts. By this means he aims to characterize certain ways of moving from one thought to another or of continuing A by proceeding to B that are made possible or facilitated by concepts at our disposal. These transitions, however, are essential tools for structuring our way of viewing the world and will have to be taken into account every time we make an attempt at comparing our concepts and those of a foreign, or alien, community of speakers.

This is a subject to which Wittgenstein makes an inimitably brief and at the same time puzzling contribution by recommending, on the last page but two of the former second part of the *Investigations*, to compare concepts with ways of painting. If one wishes to deal with questions raised in this fashion, it may no doubt be helpful to take into account Wittgenstein's above-quoted idea that “Only the whole is the instrument, the concept”. And in order to take it into account it will surely be necessary to make good sense of his words. What is obvious is that assigning these observations to the context of questions concerning “The Inner and the Outer” — and hence to a determinate, perhaps ontologically tinged sub-area of the philosophy of mind — will not contribute to successfully dealing with the task of making sense of those words.

But there is at least one further respect in which the idea of dividing Wittgenstein's remarks into those on the philosophy of psychology, on the one hand, and those on other sorts of subject, on the other, turns out to be a drawback. Let us return to the notion of pretending as well as the closely connected image of our hidden inner life and have a look at the following passage:

The inner is hidden from us means that it is hidden from us in a sense in which it is not hidden from *him*. And it is not hidden from the owner in this sense: *he utters it* and we believe the utterance under certain conditions and there is no such thing as his making a mistake here. And this asymmetry of the game is brought out by saying that the inner is hidden from someone else.

Evidently there is an aspect of the *language-game* which suggests the idea of being private or hidden — and there is also such a thing as hiding of the inner.

[...]

If I can never know what he is feeling, then neither can he pretend.

For pretending must mean, after all, getting someone else to *make a wrong guess* as to what I feel. But if he now guesses right and is certain of being right, then he knows it. For of course I can also get him to guess *right* about my feelings and not be in doubt about them. (LW II, pp. 35–6)

There are two reasons for quoting these remarks at such length and as representative of many other ones. First, the quotation shows nicely that Wittgenstein himself in explicitly dealing with questions concerning the inner and the outer will straightaway proceed to establish connections with other areas of discussion. Here, for instance, he soon leaves the philosophy of mind to consider the idea of language-games. And he continues to make connections with the notion of privacy — which is well-known from and obviously central to the argument of *Philosophical Investigations* — by underlining a certain detail, viz., the point that what is private in the relevant sense will tend to be understood as being hidden.

The second point I want to make seems particularly important to me. The quoted passage makes it perfectly clear that Wittgenstein has reasons for occupying himself with the concepts of knowledge and error that result from his dealing with questions of the relation between the inner and the outer or, rather, with questions concerning our conception of this relation. The insights into our concepts of knowledge and error gained by dealing with these questions will, at least for the time being, remain quite independent of Wittgenstein's responses to problems raised by G. E. Moore or Norman Malcolm. And here we should remember the originality of Wittgenstein's approach as opposed to the relative conventionality of the questions at the root of Moore's and Malcolm's epistemological problems and their efforts to refute sceptical objections to our knowledge claims.

Now, while it is not possible to specify the exact date of the remarks quoted just now, they can be found in a manuscript of which we can be certain that it was not written before 1949 and which contains a fair number of remarks whose wording and substance will remind readers of the material published in *On Certainty*. It is against this background that it appears quite natural to ask the following question: in which way should we read this material if the published book did not begin (as it does now) with a quotation from Moore and a corresponding note by the editors? — if, for example, it began with the manuscript on the inner and the outer just quoted from or with an extract from this manuscript focusing on the subject matter of *On Certainty*?

Accordingly, readers of Wittgenstein's latest writings should never lose sight of the fact that the organisation of the published volumes we are familiar with mirrors the state of philological knowledge and the philosophical preferences of the editors to an extent which, at least in some cases, may justify our speaking of arbitrariness or whim. The orientation of the book *On Certainty* towards epistemological and specifically Moorean questions has already been mentioned, and we have also observed that the second volume of *Last Writings on the Philosophy of Psychology* contains a fair number of remarks that might find equally



(or better) accommodation in the volume *On Certainty*. What may come as a surprise is that analogous observations apply to the book *Remarks on Colour*. While it is true that its first part derives from a selection made by Wittgenstein himself and that, at a pinch, one might consider the title “On Colour” appropriate to this first part, the much longer part III would surely find a more adequate home in the same volume as *On Certainty*. In other words, it would have been much wiser if, half a century ago, the editors had made up their minds to publish Wittgenstein's late manuscripts in their probable chronological order without making any attempt at dividing this material into thematic groups that have little or nothing to do with Wittgenstein's own understanding of his way of doing philosophy.

But even among the remarks selected by Wittgenstein himself and published as part I of *Remarks on Colour* one will find illuminating observations on the topic I should like to make the centre of the rest of my comments. This is a topic which is commonly associated with the book *On Certainty*, but here I should like to remind you that, in one form or another, it has been a lifelong occupation of Wittgenstein's. In the following passage (which is rarely quoted) he writes:

Sentences are often used on the borderline between logic and the empirical, so that their meaning changes back and forth [[sense crosses back and forth]] and they count now as expressions of norms, now as expressions of experience.

(For it is certainly not an accompanying mental phenomenon – this is how we imagine ‘thoughts’ – but the use, which distinguishes the logical proposition from the empirical one [[a logical sentence from an empirical one]].) (RC I, §32, cf. RC III, §19)

This quotation is remarkable in several respects:

- (1) It should be observed that here Wittgenstein speaks of the *use* of *Sätze* — sentences —, which is something he does comparatively rarely. In emphasizing this point I am not concerned about a question of idiom; what concerns me is giving a correct description. For what we do in most ordinary cases is that we use words to form sentences in order to communicate something by means of them or to fulfil some other purpose. To be sure, as speakers of a language we do have a repertoire of words that we can make use of, but we do not in the same sense have command of a repertoire of sentences. Some speakers — such as politicians or other people compelled to make numerous public statements — will indeed have a repertoire of set phrases and oft-repeated sentences. But there are many situations in life where even these speakers will not help themselves from this supply but express themselves more or less spontaneously (as we like to put it). Accordingly, we should understand Wittgenstein's talk about “using” sentences as an abbreviation for



the cumbersome but presumably more accurate phrase “use of words to form and utter sentences”. The idea of a repertoire of sentences, however, is one to which I shall come back.

- (2) If you think about it, you will get the impression that, in talking about a sense’s moving back and fro across the border between the realms of the empirical and the logical, Wittgenstein steps forward on difficult, perhaps even dangerous terrain. If sense crosses over first to this and then to that side, it will have to be one and the same sense that can be found first here and then there. But if the sense remains the same, the sentence will have to be understood in the same way on both occasions, that is, as saying the same thing both times. But in spite of this Wittgenstein writes that the sentence in question is on one occasion to be regarded as expressing a norm and on another occasion as expressing something empirical). But isn’t there an enormous difference between an empirical sentence and a normative one such as an instruction or a rule? And if there really is such a difference, doesn’t it concern the content of the relevant sentence, and hence its sense — that is, what it amounts to?
- (3) To discover an answer to this question we should have a close look at the end of the quoted remark. It is here that Wittgenstein emphasizes that it is its employment (*Verwendung*) which distinguishes a logical sentence from an empirical one. In the original, unrevised version of this remark Wittgenstein adds that this employment is something “surrounding” or “enveloping” a thought or a sentence. This addition clarifies what may well be obvious anyway, viz. that the way a sentence is to be understood — either as a norm or as an empirical statement — will depend on the context of utterance. This would be compatible with the idea that a sentence, while retaining the same sense, can none the less play different roles. Just think of an example like the following claim discussed in Wittgenstein’s *Remarks on Colour*: “There is no such thing as transparent-white glass.” One person may understand this as an empirical report, another one may think that it is a rule guiding our use of words, and a third person may interpret it as an analytic or a conceptual truth. Which of them is right is a question whose answer will depend on the relevant circumstances, that is, on features of the situation of utterance. But even if we accept it as true that only one of them can be right, this would not necessarily mean that each of the three hearers really *means* something different by these words, that is, assigns a *different* sense to the sentence.

As far as Wittgenstein’s last writings are concerned, his probably best-known examination of the issue just broached can be found in the context of his development of the metaphor of the riverbed, which, in *On Certainty*, serves him to elu-

cidate the notion of a world-picture. This is not the place to quote the whole sequence of relevant remarks, nor will it be to the point to explain its whole significance. Here, I shall content myself with quoting three paragraphs, and even these few paragraphs will only be discussed to the extent necessary to throw some light on our previous quotation. Wittgenstein writes as follows:

It might be imagined that some propositions (*Sätze*), of the form of empirical propositions, were hardened and functioned as channels for such empirical propositions as were not hardened but fluid; and that this relation altered with time, in that fluid propositions hardened, and hard ones became fluid.

The mythology may change back into a state of flux, the river-bed of thoughts may shift. But I distinguish between the movement of the waters on the river-bed and the shift of the bed itself; though there is not a sharp division of the one from the other.

But if someone were to say "So logic too is an empirical science" he would be wrong. Yet this is right: the same proposition may get treated at one time as something to test by experience, at another as a rule of testing. (OC, §§96–8)

Like most, or possibly all, the well-known images to be found in *On Certainty*, this metaphor of the riverbed is a metaphor involving movement and being at rest: the unyielding, immobile elements at rest make it possible for the mobile, fluid (or flowing?) elements to move. Here, the hard elements correspond to "logical" sentences or "rules" (as Wittgenstein sometimes calls them), while the fluid, flowing, mobile elements correspond to empirical sentences.

A closer look will reveal a number of differences between the quotation taken from *Remarks on Colour* and the metaphor of the riverbed from *On Certainty*: the metaphor does not involve anything which would really fit the borderline mentioned in the earlier quotation. Of course, one might want to say that the bed itself — that is, the hard part — forms a kind of barrier obstructing the flow of the water, the fluid element. But in reality it only forms a barrier to the extent it determines the direction of the flow of the river. It doesn't prevent access — on the contrary, as Wittgenstein points out, there are transitory states and boundaries are fluid. For only certain parts of the bank of the river are unyielding; they consist "of hard rock" and permit no water to penetrate. Other parts, however, can blend with water, they let it pass into the harder substances which are then carried off until they settle at a different place, where they may stay for a while. In the earlier quotation, the freedom of movement consisted in this: that the sense of a sentence was allowed to cross over from one side of the border to the other one. The riverbed, however, is in unrestricted flux, for even its most rigid elements are not as resistant as a borderline drawn in the abstract. To be sure, there is (as Wittgenstein underlines) a distinction between the everyday traffic of fluid waters (or ordinary empirical statements) and the supporting, very slowly

shifting movement of the riverbed (that is, the logical scaffolding). On the other hand, there is no “sharp division” between them — an expression which is varied and pointedly repeated in Wittgenstein’s manuscripts of this period. As a result of our comparison between these two passages we may hence come to the conclusion that not only the border between the two areas in question — that is, between the empirical and the logical — has turned out to be fluid. It also becomes clear that the sentences themselves are not rigid. And this means that their roles are not given once and for all: some of them are in a state of transition while other ones are moving in this direction or in that.

By employing this metaphor of the riverbed particular importance is given to an historical, developmental perspective. But would it not be possible to cut through the development to record by means of a picture of a certain moment in time which sentences belong to the set of (more or less) rigid ones and which to the set of (more or less) fluid sentences? In other words, is there not, at least in theory, a possibility of specifying which sentences are subject to empirical examination and which function as rules guiding or governing such an examination?

At first blush these questions may sound quite reasonable. But at the same time these questions express a certain inclination which goes against the grain of Wittgenstein’s own remarks. The inclination I mean comes to the fore in a popular interpretation of another famous metaphor from Wittgenstein’s *On Certainty*. Here I am alluding to the idea of *hinges*, and hence to yet another metaphor for movement and staying at rest. For it is the fixed and immobile hinge which makes it possible for the door to turn freely. This thought is articulated in the following remarks:

[...] the *questions* that we raise and our *doubts* depend on the fact that some propositions are exempt from doubt, are as it were like hinges on which those turn.

[...] If I want the door to turn, the hinges must stay put. (OC, §§341, 343)

Now, it seems that these remarks are often read as meaning that in Wittgenstein’s view there is a certain class or type of sentences which for us as speakers of a given language *L* or members of a certain culture *C* fulfil the function of hinges by means of which all the other sentences are given the possibility to turn. Basically, we should then be dealing with a variant of the classical picture of foundations. For the hinges would be analogous to foundations supporting certain sentences and thus allowing them to evolve and serve certain functions.

In some areas of recent discussion the metaphor has acquired independent suggestive power which has led a number of authors to speak of “hinge propositions” or “hinges” *simpliciter*. No doubt this way of talking, which implies the existence of determinable sets of such propositions, suits the general conception

of some of these authors, but it is *not* a conception favoured by Wittgenstein in this or any other passage. For Wittgenstein does not say that *specific* sentences belonging to a given repertoire need to stand firm for our everyday sentences to be able to move freely and thus to do their ordinary work. What he does say is that there have to be *some* sentences whose standing firm allows other sentences to move. These immobile sentences may vary from speaker to speaker and from one situation of utterance to another. In addition, there are a few passages where Wittgenstein indicates that a hinge may turn out to be something like a constructed or an inferred axis.

At any rate, it is important to understand that from Wittgenstein's perspective there is no determinate list — no repertoire — of basic or foundational propositions. This claim, however, is quite compatible with the requirement that some sentences (such as "Water boils at 100°") play a role analogous to articles of faith and are learned like the catechism (see the obvious "creed" in OC §234). And it is also compatible with Wittgenstein's claim that equations and other *mathematische Sätze* are, as he says, placed in the archives, that they are as it were "fossilized" and given the official "stamp of incontestability" (OC §§657, 655). Sentences of these types do belong to a given repertoire which everyone may draw on and which everyone should conform to. But many sentences whose peculiar role in an individual's life attracts Wittgenstein's attention don't in the least belong to this category. Just think of examples like "I know that my name is Joachim Schulte", "These last weeks I have had a shower every day" and so on. From *my* point of view, such sentences can throw an interesting light on *my* world picture, on *my* system of investigation and judgement as well as on *my* understanding of this system. But it should be obvious that sentences of these kinds are in no way general, they are not mandatory, and they belong to no repertoire which all members of a culture or a linguistic community are supposed to swear by.

On the contrary, the notion of such a repertoire may tend to prevent us from really appreciating two ideas that are particularly characteristic of the later Wittgenstein. The first idea is this: that, on the one hand, a distinction between logical and empirical sentences will involve the notion of a sharp boundary between them. But, on the other hand, such a distinction will have to work in a context of tensions and conflicting near-certainties which, from an historical point of view as well as regarding interpersonal applicability, can rely neither on any clear nor on any generally acknowledged boundaries. Wittgenstein himself summarizes this as follows. He writes that

there is no sharp boundary between methodological propositions and propositions within a method. [...] The lack of sharpness is that of the boundary between *rule* and empirical proposition. (OC §§318–19)

The second idea comes to the fore in the immediately following remark which says that we should remember that the notion of a sentence itself is not a sharp one. Well, we *want* to say that one and the same sentence can play different — as it were, *logically*, different — roles. At the same time, however, we are tempted to describe this sort of difference in a way which is not in harmony with our conception of the identity of the sentence. As Wittgenstein suggests, this is a situation which we shall have to live with. For it is *our* situation. This is, as I have implied, an insight foreshadowed in earlier writings by Wittgenstein. But nowhere is it expressed as strikingly and with such metaphorical succinctness as in his last manuscripts.

## Bibliography

- Bouwsma, O. K. (1986): *Wittgenstein: Conversations 1949–1951*. Edited by J. L. Craft and Ronald E. Hustwit. Indianapolis: Hackett.
- McGuinness, Brian (ed.) (2012): *Wittgenstein in Cambridge: Letters and Documents 1911–1951*. Revised edition. Malden, MA, and Oxford: Wiley-Blackwell. [=WiC]
- Wittgenstein, Ludwig (1974): *On Certainty*. Edited by G. E. M. Anscombe and G. H. von Wright 1969; reprinted with corrections. Oxford: Blackwell. [= OC]
- Wittgenstein, Ludwig (1977): *Remarks on Colour*. Edited by G. E. M. Anscombe. Oxford: Blackwell. [= RC]
- Wittgenstein, Ludwig (1980): *Remarks on the Philosophy of Psychology*, Volume I. Edited by G. H. von Wright and Heikki Nyman. Oxford: Blackwell.
- Wittgenstein, Ludwig (1980): *Remarks on the Philosophy of Psychology*, Volume II. Edited by G. H. von Wright and Heikki Nyman. Oxford: Blackwell.
- Wittgenstein, Ludwig (1982): *Last Writings on the Philosophy of Psychology*, Volume I. Edited by G. H. von Wright and Heikki Nyman. Oxford: Blackwell.
- Wittgenstein, Ludwig (1992): *Last Writings on the Philosophy of Psychology*, Volume II. Edited by G. H. von Wright and Heikki Nyman. Oxford: Blackwell.
- Wittgenstein, Ludwig (1998): *Culture and Value*. Third revised edition. Oxford: Blackwell.
- Wittgenstein, Ludwig (2000): *Wittgenstein's Nachlass: The Bergen Electronic Edition*. Oxford: Oxford University Press.
- Wittgenstein, Ludwig (2009): *Philosophical Investigations*. Revised fourth edition by P. M. S. Hacker and Joachim Schulte. Malden, MA, and Oxford: Wiley-Blackwell.



## Metaphilosophy and Methodology



Dan Zahavi

# Analytic and Continental Philosophy: From Duality Through Plurality to (Some Kind of) Unity

**Abstract:** In the paper, I assess some of the categorical claims that over the years have been made regarding the relation between analytic and Continental philosophy. I will argue that many of these statements involve stupendous oversimplifications and overlook many significant differences. I next compare ideas found in phenomenology with ideas found in analytic philosophy of language and analytic philosophy of mind respectively and ultimately argue that the right way to overcome the analytic-Continental divide is by embracing plurality rather than unity.

## 1 Perplexing hostility

What should we make of the distinction between analytic and Continental philosophy? Although the latter term is a more recent invention, the alleged division has been around for a while. My aim in the following contribution is not to trace and review its historical origin, nor do I intend to examine to what extent debates and disputes between seminal figures like Frege and Husserl, Russell and Bergson or Carnap and Heidegger might have helped to shape and cement the division. My focus will be different and more direct. Is it at all an informative distinction? Does it label a real divide, a divide between what might have become two incommensurable conceptions of philosophy? That this is indeed a view held by some is easy to exemplify.

In *Origins of Analytical Philosophy*, Dummett refers to the divide which he argues had widened to a point where “[i]t’s no use [...] shouting across the gulf” since “it’s as if we’re working in different subjects” (Dummett 1996, p. 193). Now, Dummett regrets the divide and suggests a way of repairing it – primarily by going back to the beginning of the divergence in order to examine the partially overlapping projects of Frege and Husserl, including their shared rejection of psychologism. Others, however, have accepted the existence of the split, but have seen no particular reason to regret or amend it. According to R.M. Hare, for instance, German philosophers have created “monstrous philosophical edifices” and produced ‘verbiage’ disguised as “serious metaphysical inquiry”, and he contrasts the virtues of British philosophy, namely “clarity, relevance and brevity”, with the “ambiguities and evasions and rhetoric” that are the “mark of a philosopher who has not learnt



his craft” (Hare 1960). Jack Smart is even more condemning and declares that “I have moments of despair about philosophy when I think of how so much phenomenological and existential philosophy seems such sheer bosh that I cannot even begin to read it” (Smart 1975, p. 61). For a quite recent statement of the same sentiment, consider Quinton’s entry on Continental philosophy in *The Oxford Companion to Philosophy*. Quinton opines that there is “no perceptible convergence between the two philosophical worlds” since all varieties of Continental philosophy “rely on dramatic, even melodramatic, utterance rather than sustained rational argument”. And, as he then concludes, for analytic philosophers, it can at most be “the object of occasional startled observation, like that of a nasty motor accident viewed from a passing car” (2005, p. 172)

Such dismissive and ridiculing remarks are, however, by no means exclusive to analytic philosophers. In *Negative Dialectics*, Adorno remarked that without dialectics, arguments deteriorate into a mere technique that “is now spreading academically in the so-called ‘analytical philosophy,’ which robots can learn and copy” (2004, p. 30) and, in a newspaper article from 2013, Zabala & Davis have concluded that

Analytic philosophy is passé because its method is too conservative to transgress the pre-suppositions on which it is based. This is not only why analytic philosophy is “anal” (in the Freudian anal-retentive sense), but its conservative nature binds it to a method that has already died. (Zabala & Davis 2013)

More examples could be found, but enough has been said to show that many do indeed believe in the existence of a divide, perhaps even an unbridgeable one, between analytic philosophy and Continental philosophy. But what then does the difference amount to? Over the years, a variety of proposals have been made. The most immediate suggestion might be to look for a difference in focus. Analytic philosophy and Continental philosophy differ in virtue of the themes and topics they work on. The moment one sits down to actually list which topics are unique to each of the traditions, it becomes evident, however, that such a listing is doomed from the start. As Overgaard et al aptly put it:

Hardly any feature held to be characteristic of one of the two camps is universally shared by all who belong to the camp in question. And most features...*are* shared by philosophers from the other side of the divide. (Overgaard et al. 2013, p. 113)

But perhaps we are looking in the wrong direction. Perhaps the relevant difference is not one of topic, but rather one related to method. But, again, a moment’s reflection will show that this proposal is equally problematic. Philosophers usually considered to belong to the Continental camp do not use one

specific method (what should unite phenomenologists, critical theorists and poststructuralists?), and most contemporary analytic philosophers are methodological pluralists. They are no longer committed to the ‘linguistic turn’ and would hardly accept the suggestion that a “primary method of philosophy is the examination of the uses of words in order to disentangle conceptual confusions” (Hacker 2007, p. 133).

What else might the difference consist in? As Bernard Williams once remarked, analytical philosophy doesn’t have a distinctive subject matter, but is rather characterized by a certain style (2006, p. viii). Again, it is unclear what exactly this difference should amount to. Is the idea that philosophers belonging to the Continental camp write like novelists, whereas analytic philosophers seek to emulate the style of scientific papers? Or is the idea that the writing of analytic philosophers is distinguished by its clarity, whereas that of Continental philosophers is distinguished by its obscurity? In his book, *What is analytic philosophy?*, Glock considers and exemplifies the style of various analytic philosophers and eventually concludes: “the speech of many contemporary analytic philosophers is as plain as a baroque church and as clear as mud” (Glock 2008, p. 171). A closer look at the actual writings of many of the leading figures will quickly show that there is no easily determinable stylistic commonality. It is quite unclear what should unite the writing style of Apel, Adorno, Cassirer, Husserl and Derrida. Do Nagel and Peacocke, Dennett and McDowell, Wittgenstein and Searle or Cavell and Tye really write in the same style? The answer is obviously no.

Are there other promising proposals regarding the alleged difference? Let me just mention one more, which, although inane, certainly possesses an elegant simplicity: the difference between analytic philosophy and Continental philosophy is a difference in quality. As Leiter puts it,

It is fair to say that what gets called “analytic” philosophy is the philosophical movement most continuous with the “grand” tradition in philosophy, the tradition of Aristotle and Descartes and Hume and Kant. Only analytic philosophers aspire to the level of argumentative sophistication and philosophical depth that marks the great philosophers. (Leiter 2011)

As was quite clear from the quote by Smart given above, the intensity of such ridicule is often correlated with a lack of familiarity with the target of criticism. Perhaps a more well-informed conclusion to draw is that there really isn’t any philosophical difference to speak of – although nobody would deny the sociological and institutional persistence of the divide. As Putnam writes, “why can we not just be philosophers without an adjective?” (Putnam 1997, p. 203) This is a sympathetic response, but one, I think, that is not only misguided, but also potentially harmful. Our philosophical approach is not ahistorical, but is shaped

and formed by the tradition(s) of which we are part and, as critical thinkers, we ought to remain aware of where we come from and of how that background and framework influence the questions we engage with and the conceptual options we consider. Simply thinking of ourselves as philosophers is not going to help us remain mindful of these facts.

Let me get down to my own core proposal. As I see it, the right way to question the divide between analytic philosophy and Continental philosophy is not by denying the difference, but by multiplying it. The main reason we ought to reject any talk of a divide between analytic philosophy and Continental philosophy is that such a division is far too coarse-grained and commits the mistake of regarding both (sets of) traditions as reified monolithic entities. Consider for a moment the following list, which certainly isn't exhaustive: Philosophy of life, Neo-Kantianism, Phenomenology, Hermeneutics, Existential philosophy, Critical theory, Structuralism, Psychoanalytic theory, Post-structuralism, Deconstruction, Speculative realism. What is the commonality here? (Were one to claim that genealogy – rather than substantial agreement – is enough to constitute the unity of a tradition, we would all be Platonists). Perhaps it could be retorted that, whereas Continental philosophy is indeed a mixed bag of assorted theory formations with no clear unity, matters are different when it comes to analytic philosophy. Consider, for instance, Dummett who first describes Frege as “the fountain-head of analytical philosophy” (1978, p. 440), and who then continues:

Only with Frege was the proper object of philosophy finally established: namely, first, that the goal of philosophy is the analysis of the structure of *thought*; secondly, that the study of *thought* is to be sharply distinguished from the study of the psychological process of *thinking*; and, finally, that the only proper method for analysing thought consists in the analysis of *language*. [...] The acceptance of these three tenets is common to the entire analytical school. (Dummett 1978, p. 458)

More recently, Hacker has taken issue with Dummett's characterization of Frege's contribution to the linguistic turn – which he insists is radically mistaken (Hacker 2007, p. 134) – and has argued that we ought to reserve the term ‘analytic philosophy’ as a label for a quite specific period in the history of philosophy (centered around the cities of Cambridge, Vienna and Oxford). On Hacker's construal, the most influential version of analytic philosophy is found in the work of Oxford philosophers such as Ryle, Austin, and Strawson, who broadly agreed on three issues (Hacker 2007, p. 126–127):

- Philosophical understanding must proceed through an investigation of the use of words;

- Metaphysics understood as a philosophical investigation into the objective, language-independent nature of the world is an illusion;
- Philosophy is not continuous with but altogether distinct from science.

Hacker's definition of analytic philosophy is fairly precise, but it is also a quite more narrow and circumscribed use of the term, and it certainly doesn't capture the way the term is used when referring to the current divide between analytic philosophy and Continental philosophy. To put it differently, quite a number of contemporary philosophers who call themselves analytic philosophers would reject the linguistic turn, happily engage in metaphysics, and even argue that philosophy is continuous with science and ought to be thoroughly naturalized.

There is, in short, plenty of diversity. Recognizing this diversity, recognizing that, just as Althusser, Scheler and Deleuze differ, so do Austin, Korsgaard and Churchland, should not only caution us against postulating a unity where there isn't any, but might also and more importantly allow us to discover unity where we didn't expect it. We should search for overlaps that are not merely within, but also between the traditions. Depending on the kind of philosophical work one is engaged in, one might turn out to have more in common with people working in "the" other (set of) tradition(s) than with people working in one's own.

The moment we abandon the attempt to adopt a bird's eye perspective from which we can compare and contrast analytic and Continental philosophy as a whole and instead start to engage with the actual texts themselves, interesting and unexpected convergences will be revealed. Compare for example Husserl's early work on intentionality with more recent contributions by Searle, Strawson, Kriegel and Crane. Take Heidegger's and Merleau-Ponty's exploration of our being-in-the-world and compare that to the recent debate on 4E cognition found in the work of Clark, Noë, Rowlands, Menary and Hutto. Or consider the analyses of shared affects and we-intentionality found in Scheler, Walther, Stein and Schütz and compare that to contributions to philosophy of action and collective intentionality made by authors such as Bratman, Gilbert and Tuomela. Sure, there are important differences, but to claim that any attempt at dialogue is doomed from the start since people will necessarily be talking at cross purposes due to the endorsement of too divergent methodological and metaphysical commitments is surely mistaken.

Let me in the following discuss two concrete examples in somewhat more detail. I will first consider the relation between transcendental phenomenology and classical philosophy of language, and then compare ideas found in more recent philosophy of mind with themes found in phenomenology.

## 2 Phenomenology and philosophy of language

At first sight, the difference between phenomenology and analytic philosophy of language seems obvious. Whereas the latter is precisely concerned with language, the former is typically taken to investigate structures of experience. Supposedly, this difference was also highlighted at one of the rare meetings where proponents of both traditions met.

At the famous 1958 Royaumont meeting, Gilbert Ryle, J.L. Austin, W.V.O. Quine, Bernard Williams and Peter Strawson met with phenomenological notabilities such as Jean Wahl, Maurice Merleau-Ponty and H.L. van Breda.<sup>1</sup> The meeting was no success. As Charles Taylor, who was also in attendance, subsequently reported, it was “a dialogue that didn’t come off [...]. [F]ew left it much wiser than they came – at least as far as the subject of the conference was concerned” (Taylor 1964, p. 132–132). At one point during the meeting, P.F. Strawson expounded on his own view of what the principal task of philosophy is, and insisted that language use is the only experimental datum philosophers possess if they want to inquire into the way in which our thoughts and concepts work (Strawson 1992, p. 324). This proposal was resisted by Van Breda who interrupted Strawson with the outburst “But you have to distinguish what you are doing from what the philologist does!” (quoted in Strawson 1992, p. 327) Van Breda’s criticism was certainly based on a misunderstanding. Strawson’s philosophy is neither to be conflated with philology nor sociolinguistics. But it is a revealing misinterpretation, especially since it mirrors an often-repeated criticism of phenomenology, according to which phenomenological investigations of experience employ introspection and suffer from the same kind of fatal weakness as introspectionism. Why are both criticisms mistaken? Because they both overlook the transcendental agenda of the targeted positions.

As Merleau-Ponty points out in *Phénoménologie de la perception*, phenomenology is distinguished in each of its characteristics from introspective psychology, and this is because they differ in principle. Whereas the introspective psychologist considers consciousness one sector of being, and tries to investigate this sector in the same way the physicist tries to investigate his, the phenomenologist realizes that an investigation of consciousness cannot take place as long as the absolute existence of the world is left unquestioned. Consciousness cannot be analyzed properly without leading us beyond common-sense assumptions and towards a transcendental clarification of the constitution of the world (Merleau-Ponty 2012, p. 60–61). One way to understand Merleau-Ponty’s

---

<sup>1</sup> I am here drawing on an excellent contribution by Overgaard (2010).

point is to see him insisting that the critical stance proper to philosophy necessitates a move away from a straightforward metaphysical or empirical investigation of objects to an investigation of the very framework of meaning and intelligibility that makes any such straightforward investigation possible in the first place. Rather than departing from the objective world as a given, phenomenology asks how something like objectivity is possible. How is objectivity constituted? How is it that the world can be manifested or revealed to us in the first place? This was also a question of central concern to Husserl, whose claim was that, if we really want to understand meaning, knowledge, truth and objectivity, we have to investigate the intentional structures of the cognizing and acting subjects. As he declared in *Krisis*, thereby summarizing his lifelong endeavor:

The first breakthrough of this universal a priori of correlation between experienced object and manners of givenness (which occurred during work on my *Logical Investigations* around 1898) affected me so deeply that my whole subsequent life-work has been dominated by the task of systematically elaborating on this a priori of correlation. (Husserl 1970, p. 166)

It is not difficult to see some affinities between such a view and one that argues that philosophy offers a kind of understanding that is distinctively different from the one offered by the sciences. Philosophical questions are second-order questions, reflective questions about the concepts used in, for example, scientific fields; they are not first-order questions within a particular scientific field. Philosophy is not concerned with matters of fact, but with matters of meaning. It is concerned with logical possibilities rather than with empirical actualities. Its province is not the domain of empirical truth or falsehood, but the domain of sense and nonsense. To put it differently, philosophy clarifies what does and does not make sense. It investigates and describes the bounds of sense: that is, the limits of what can coherently be thought and said. Conceptual questions antecede matters of truth and falsehood. They are presupposed by any scientific investigation, and any lack of clarity regarding the relevant concepts will be reflected in a corresponding lack of clarity in the questions posed, and hence in the design of the experiments intended to answer them (Bennett & Hacker 2003, p. 2, 399, 402). As Hacker writes:

Philosophy is not an extension of science. It is not a kind of conceptual scullery maid for the sciences, as Locke supposed. Nor is it superior to the sciences – a super-science of all possible worlds, to be investigated by means of “thought-experiments” from the comfort of the armchair, as contemporary revisionists suppose. (Thought-experiments are no more experiments than monopoly-money is money.) It is, as Kant intimated, the *Tribunal of Sense*. So: back to the linguistic turn. (Hacker 2007, p. 139)

Perhaps some would at this stage object that the focus on language surely points to one significant difference between phenomenology and analytic philosophy. That is indeed an important difference, but, then again, we shouldn't be so naïve as to think that phenomenology has nothing to say about language. That Heidegger (and Gadamer) is concerned with language is after all common knowledge, but so was Husserl, as the following quote shows:

A communal life of humans becomes possible as life of a linguistic community, which is of a completely different kind than the communal life of animals. *The homeworld of humans*, which is the fundamental element for the structure of the objective world for <them> (...) is essentially determined by language (Husserl 1973, p. 224–5).

I am far from suggesting that there is no significant difference between the respective approaches. My point is rather that the moment we leave the simplistic caricatures behind, and start to engage with the actual writings and sayings of the protagonists, we will start to discover unforeseen similarities as well as unexpected differences. Consider, for instance, Hacker's view on the relationship between conceptual and empirical issues. For him, philosophy is of much greater importance to science than vice versa. While philosophers can clarify the concepts used in science, and thereby offer an immense service to science, it is a mistake to think that science could have much of an impact on philosophy. In fact, Hacker even considers the supposition that scientific evidence may contravene a philosophical analysis ridiculous (Bennett & Hacker 2003, p. 404). In his view, we should not commit the mistake of confusing metaphysical or epistemological theories with empirical claims which can be corroborated by some *experimentum crucis*. Thus, the relation between philosophy and empirical science is a one-way enterprise. It is an application of ready-made concepts. There is no reciprocity, and there is no feedback. The application does not lead to a modification of the original analysis.

This is certainly one type of response to the challenge posed to philosophy by naturalism. It is perhaps not entirely without reason that the style of analytical philosophy defended by Hacker has occasionally been accused of promoting a kind of semantic inertia and conceptual conservatism. It is not obvious, however, that phenomenology would respond in the same way as Hacker. To let an examination of ordinary language-use be our primary, if not exclusive, guide to a philosophical investigation is quite restrictive and seems to underestimate the extent to which ordinary language might simply reflect common-sense metaphysics. It impedes concrete phenomenological analyses that might reveal features, aspects and dimensions that are not simply available to any reflection on com-

mon sense (consider for instance Husserl's investigations of the structures of time-consciousness or pictorial consciousness).

I have elsewhere discussed various proposals regarding a naturalisation of phenomenology and also highlighted the extent to which transcendental phenomenology differs from Kantian transcendental philosophy (Zahavi 2010, 2013, 2015). So let me not repeat myself *in extenso* here, but instead simply make a brief reference to Merleau-Ponty, which can highlight one significant difference between his conception of philosophy and Hacker's. In his first major work, *The Structure of Behavior*, Merleau-Ponty engages with such diverse authors as Pavlov, Freud, Koffka, Piaget, Watson, and Wallon. The last sub-chapter of the book carries the heading "Is There Not a Truth in Naturalism?" It includes a criticism of Kantian transcendental philosophy and, on the very final page of the book, Merleau-Ponty calls for a redefinition of transcendental philosophy that pays heed to the real world (Merleau-Ponty 1963, p. 224). Thus, rather than making us choose between either an external scientific explanation or an internal phenomenological reflection, a choice which would rip asunder the living relation between consciousness and nature, Merleau-Ponty asks us to reconsider the very opposition, and to search for a dimension that is beyond both objectivism and subjectivism.

What is interesting and important is that Merleau-Ponty didn't conceive of the relation between transcendental phenomenology and empirical science as a question of how to apply already established phenomenological insights to empirical issues. It wasn't merely a question of how phenomenology might constrain positive science. On the contrary, Merleau-Ponty's idea was that phenomenology itself can be changed and modified through its dialogue with the empirical disciplines. In fact, it needs this confrontation if it is to develop in the right way. And Merleau-Ponty held on to this view without thereby reducing phenomenology to merely yet another empirical science, without thereby dismissing its transcendental philosophical character.

### 3 Phenomenology and philosophy of mind

An increasing number of analytic philosophers of mind now hold the view that a careful study of the first-person perspective is indispensable for a philosophical investigation of consciousness. More specifically, a number of authors have made significant strides on the relation between phenomenal consciousness, self-experience, and selfhood. That such a development offers new possibilities for dialogue should be evident. Let me briefly present a few ideas found in the work of Galen Strawson and Uriah Kriegel.



Over the years, Strawson has argued that if we wish to answer the metaphysical question concerning whether or not the self is real, we will first need to know what a self is supposed to be. In order to establish that, our best chance is to look at self-experience, since self-experience is what gives rise to the question in the first place by giving us a vivid sense that there is something like a self. Thus, as Strawson readily concedes, the metaphysical investigation of the self is subordinate to the phenomenological one. The latter places constraints on the former: nothing can count as a self unless it possesses those properties attributed to the self by some genuine form of self-experience (Strawson 2000, p. 40). In a subsequent move, Strawson has argued that a phenomenological investigation can proceed in several ways. One possibility is to investigate what ordinary human self-experience involves; another is to investigate the minimal form of self-experience. What is the least you can make do with to still call something a(n experience of) self? Strawson is primarily interested in the latter task, and he defends the view that any experience involves an experiencer, i.e., a subject of experience. Experience necessarily involves what-it-is-likeness, and experiential what-it-is-likeness is necessarily what-it-is-likeness for someone (Strawson 2009, p. 271). In short, experience is experiencing, and experiencing involves a subject, just as a branch-bending involves a branch (Strawson 2011, p. 260). Importantly, Strawson doesn't simply take this to be a conceptual and metaphysical claim. It is also an experiential or phenomenological claim: the subject of experience is something that is essentially present and alive in conscious experience (Strawson 2009, p. 362). Moreover, we are dealing with a quite minimal notion. A subject of experience isn't something grand. It is minimal in the sense that it is what is left when everything else except the being of experience is stripped away (Strawson 2011, p. 254). In fact, it is something of such a kind that it is true to say that there must be a subject of experience wherever there is experience, even in the case of mice or spiders, or sea snails—simply because selves are just subjects of experience and experience is essentially experience-for (Strawson 2009, p. 276, 401).

Along somewhat related lines, Kriegel has argued that phenomenal character involves both qualitative character, e.g., the bluish component of 'x', and subjective character, i.e., its for-me component (Kriegel 2009, p. 8). Kriegel further describes the subjective character as that which remains invariant across all phenomenal characters, and argues that, while a phenomenally conscious state's qualitative character is what makes it the phenomenally conscious state it is, its subjective character is what makes it a phenomenally conscious state at all (Kriegel 2009, p. 2, 58). More specifically, Kriegel has argued that the subjective character (and experiential for-me-ness) amounts to a special kind of self-consciousness. On his account, we need to distinguish two types of self-consciousness, *transitive* and *intransitive* self-consciousness. Whereas transitive

self-consciousness designates the case where a subject is self-conscious of her thought that *p* (or of her perception of *x*), intransitive self-consciousness can be captured by saying that the subject is self-consciously thinking that *p* (or perceiving *x*). What is the difference between these two types of self-consciousness? Kriegel initially lists four differences, and claims that, whereas the first type is introspective, rare, voluntary, and effortful, the second is none of these. However, he then also points to another crucial difference: whereas transitive self-consciousness is a second-order state that is numerically distinct from its object, namely the respective first-order state, intransitive self-consciousness is a property of the first-order state itself (Kriegel 2003, p. 104–105). Moreover, insofar as there would not be anything it is like for a subject to be in a mental state she is unaware of being in, intransitive self-consciousness must be considered a necessary condition for phenomenal consciousness (Kriegel 2003, p. 106).

Kriegel's proposal bears an obvious similarity to the phenomenological distinction between and discussion of pre-reflective and reflective self-consciousness, just as Strawson's reference to a minimal self resonates well with ideas found in phenomenology. Consider for example the following quote from Sartre: "pre-reflective consciousness is self-consciousness. It is this same notion of *self* which must be studied, for it defines the very being of consciousness" (Sartre 2003, p. 100). The basic idea defended by several phenomenologists is that consciousness is characterized by a basic dimension of selfhood precisely because of its ubiquitous self-consciousness (cf. Zahavi 1999, 2005, 2014). To phrase it in more contemporary terms, it has been proposed that phenomenally conscious episodes, episodes characterized by a subjective what-it-is-likeness, are not merely episodes that happen to take place *in a subject*, regardless of whether or not the subject is aware of it. Rather, such episodes are necessarily pre-reflectively self-conscious in the weak sense that they are like something *for the subject*. Indeed, for every possible experience we might have, each of us can say that, regardless of whatever it is precisely like for me to have this experience, it is *for me* that it is like that to have the experience. To that extent, what-it-is-like-ness is properly speaking what-it-is-like-*for-me*-ness. This *for-me*-ness does not denote a specific experiential content; rather, it refers to the first-personal presence of experience. It refers to the fact that the experiences I am living through present themselves differently to me than to anybody else. When I have experiences, I have them *minely*, so to speak. The claim is that this first-personal presence, this *for-me*-ness, amounts to a primitive and minimal form of selfhood. As Henry formulates it, the most basic form of selfhood is the one constituted by the very self-manifestation of experience (Henry 1963, p. 581; 1965, p. 53).

Given this manifest overlap in terms of topics and arguments, it is increasingly hard to take the verdict seriously that analytic philosophy and phenomen-

ology constitute two distinct and incommensurable conceptions of philosophy.<sup>2</sup> All of this is not to deny that one cannot also find significant and interesting differences. Kriegel has, for instance, insisted that intransient self-consciousness is a form of object-consciousness, whereas Strawson has defended the claim that the self is not only an object, but a thing as physical as a cow (Strawson 1997, p. 425). Both claims are in conflict with views espoused by phenomenologists, who would typically insist that the self is first and foremost a subject, rather than an object, of experience, and that object-consciousness is singularly unsuited as a model for self-consciousness proper, since it necessarily entails a distinction between that which appears and that to whom it appears, between the object and the subject of experience. Sometimes disagreements are more terminological than substantial, however. This might, at least partially, be the case here. Strawson has, for instance, argued that for something to be an object simply means that it is a strong unity (Strawson 2009, p. 298), whereas Kriegel has admitted that the relationship between inner awareness and what it is aware of is far more intimate than the standard relationship between a representation and what it represents (Kriegel 2009, p. 107–8), and that it therefore amounts to a very unusual form of object-consciousness. The question is obviously whether the difference between a highly unusual form of object-consciousness and a non-objectifying form of self-consciousness is all that substantial.

## 4 Conclusion

As I have tried to show in the preceding, it is possible to locate some significant and fertile overlaps between phenomenology and various (rather different) trends in analytic philosophy. One interesting, and perhaps slightly surprising, outcome of this comparison is that phenomenologists might have slightly mixed feelings about the recent turn in analytic philosophy away from philosophy of language and towards philosophy of mind. With few exceptions, the recent analytic interest in the study of consciousness has gone hand in hand with a strong commitment to naturalism. Whereas classical analytical philosophy of language might have been less interested in consciousness, its basic conception of language as a framework of intelligibility could easily be given a transcendental philosophical twist. Rather than engaging in first-order claims about the nature of things (which it left to various scientific disciplines), it concerned

---

<sup>2</sup> Such a verdict is also countered by the existence of co-authored publications (cf. Zahavi & Kriegel 2016).

itself with the conceptual preconditions for any such empirical inquiries. The situation is now more or less reversed by those philosophers of mind who consider their own work to be more or less continuous with the natural sciences. For phenomenology, however, it is important to retain both aspects: a concern with subjectivity and experience as well as the transcendental perspective.

Let me sum up: it is a mistake to carve up the philosophical landscape into two distinct (and incommensurable) traditions. The mistake is both one of oversimplification and reification. There are far more than two traditions (let us also not forget the existence of Asian philosophical traditions) and, when it comes to analytical philosophy and Continental philosophy neither (set of) tradition(s) is monolithic. Acknowledging the diversity allows us to recognize the presence of unexpected similarities as well as fruitful and productive differences. Despite the undeniable sociological and institutional reality of the divide, however, the future looks more promising. An increasing number of philosophers are now active bridge-builders. They work in and with different traditions and are actively pursuing philosophical insights wherever they are to be found.

## Bibliography

- Adorno, Theodor W. (2004): *Negative Dialectics*. Trans. E.B. Ashton. London: Routledge.
- Bennett, Maxwell R. & Hacker, Peter M. S. (2003): *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.
- Dummett, Michael (1978): *Truth and Other Enigmas*. Cambridge, MA.: Harvard University Press.
- Dummett, Michael (1996). *Origins of Analytical Philosophy*. Cambridge, MA.: Harvard University Press.
- Glock, Hans-Johann (2008): *What is Analytic Philosophy?* Cambridge: Cambridge University Press.
- Hacker, Peter (2007): "Analytic philosophy: Beyond the linguistic turn and back again". In: Michael Beaney (ed.): *The Analytic Turn: Analysis in Early Analytic Philosophy and Phenomenology*. London. Routledge, pp. 125–141.
- Hare, R.M. (1960): "A school for philosophers". In: *Ratio* 2/2, pp. 107–120.
- Henry, Michel (1963): *L'essence de la manifestation*. Paris: PUF.
- Henry, Michel (1965): *Philosophie et phénoménologie du corps*. Paris: PUF.
- Husserl, Edmund (1970): *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy*. Trans. D. Carr. Evanston, IL: Northwestern University Press.
- Husserl, Edmund (1973): *Zur Phänomenologie der Intersubjektivität III. Texte aus dem Nachlass. Dritter Teil.1929–35*. Ed. I. Kern. Husserliana 15. Den Haag: Martinus Nijhoff.
- Kriegel, Uriah (2003): "Consciousness as Intransitive Self-Consciousness: Two Views and an Argument". In: *Canadian Journal of Philosophy* 33(1), pp. 103–132.

- Kriegel, Uriah (2009): *Subjective Consciousness: A Self-Representational Theory*. Oxford and New York: Oxford University Press.
- Leiter, Brian (2011): “‘Analytic’ and ‘Continental’ Philosophy”. In: *The Philosophical Gourmet Report*. <http://www.philosophicalgourmet.com/analytic.asp>. Visited on March 25, 2015.
- Merleau-Ponty, Maurice (1963): *The structure of behavior*. Trans. A.L. Fisher. Pittsburgh: Duquesne University Press.
- Merleau-Ponty, Maurice (2012): *Phenomenology of Perception*. Trans. D. A. Landes. London: Routledge.
- Overgaard, Søren (2010): “Royaumont Revisited”. In: *British Journal for the History of Philosophy* 18, pp. 899–924.
- Overgaard, Søren, Gilbert, Paul & Burwood, Steven (2013): *An Introduction to Metaphilosophy*. Cambridge: Cambridge University Press.
- Putnam, Hilary (1997): “A Half Century of Philosophy”. In: *Daedalus* 12, pp. 175–208.
- Quinton, Anthony (2005): “Continental Philosophy”. In: Ted Honderich (ed.): *The Oxford Companion to Philosophy*. Oxford: Oxford University Press.
- Sartre, Jean-Paul (2003): *Being and Nothingness: An Essay in Phenomenological Ontology*. Trans. H. E. Barnes. London and New York: Routledge.
- Smart, J. J. C. (1975): “My Semantic Ascents and Descents”. In: C.J. Bontempo & S. J. Odell (eds.): *The Owl of Minerva: Philosophers on Philosophy*. New York: McGraw-Hill, pp. 57–72.
- Strawson, Peter (1992): “Discussion of Strawson’s ‘Analysis, Science, and Metaphysics’”. In: Richard M. Rorty (ed.): *The Linguistic Turn: Essays in Philosophical Method*. Chicago: University of Chicago Press, pp. 321–330.
- Strawson, Galen (1997): “The self”. In: *Journal of Consciousness Studies* 4/5–6, pp. 405–28.
- Strawson, Galen (2000): “The Phenomenology and Ontology of the Self”. In: Dan Zahavi (ed.): *Exploring the Self*. Amsterdam: John Benjamins, pp. 39–54.
- Strawson, Galen (2009): *Selves: An Essay in Revisionary Metaphysics*. Oxford: Oxford University Press.
- Strawson, Galen (2011): “The Minimal Subject”. In: Shaun Gallagher (ed.), *The Oxford Handbook of the Self*. Oxford: Oxford University Press, pp. 253–278.
- Taylor, Charles (1964): “Review of La Philosophie Analytique”. In: *The Philosophical Review* 73, pp. 132–5.
- Williams, Bernard (2006): *Ethics and the Limits of Philosophy*. London: Routledge.
- Zabala, Santiago & Davis, Creston (2013): “Which philosophy is dead?”. In: *Aljazeera*. <http://www.aljazeera.com/indepth/opinion/2013/06/201361082357860647.html>, visited March 25, 2015.
- Zahavi, Dan (1999): *Self-Awareness and Alterity: A Phenomenological Investigation*. Evanston: Northwestern University Press.
- Zahavi, Dan (2005): *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, MA: The MIT Press.
- Zahavi, Dan (2010): “Naturalized Phenomenology”. In: Shaun Gallagher & Daniel Schmicking (eds.): *Handbook of Phenomenology and Cognitive Science*. Dordrecht: Springer, pp. 2–19.
- Zahavi, Dan (2013): “Naturalized Phenomenology: A Desideratum or a Category Mistake?”. In: *Royal Institute of Philosophy Supplement* 72, pp. 23–42.

- Zahavi, Dan (2014): *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford: Oxford University Press.
- Zahavi, Dan (2015): "Husserl and the Transcendental". In: Sebastian Gardner & Matthew Grist (eds.): *The Transcendental Turn*. Oxford: Oxford University Press, pp.228–243.
- Zahavi, Dan & Kriegel, Uriah (2016): "For-me-ness: What It Is and What It Is Not". In: Daniel O. Dahlstrom, Andreas Elpidorou & Walter Hopp (eds.), *Philosophy of Mind and Phenomenology: Conceptual and Empirical Approaches*. London: Routledge, pp. 36–53.



Charles Siewert

## For Analytic Phenomenology

**Abstract:** This chapter explains and defends the basics of an approach to the philosophy of consciousness here labeled “analytic phenomenology.” Its most salient feature is its reliance on critical first-person reflection to address theoretical issues concerning the character of experience. The defense proceeds principally via responses to two challenges to it implicit in recent philosophy. The first is a radically third-person approach to the study of consciousness advocated by Daniel Dennett, which he calls “heterophenomenology.” The second is a skepticism about reliance on introspective judgment put forward by Eric Schwitzgebel. The argument focuses on these authors’ treatment of first-person judgment regarding the richness or detail of ordinary visual spatial experience. While they purport to undermine a basic use of first-person judgment in trying to determine facts about such experience, their practice implicitly vindicates a critical reliance on it, and so illustrates and supports the practice of analytic phenomenology.

The philosophical project I will defend here I call “analytic phenomenology.” This label may suggest a union of two philosophical orientations usually seen in opposition – “analytic” and “continental” – insofar as phenomenology has been identified with the latter. My main interest, however, lies not in this notorious contrast, but simply in putting forward a worthwhile approach to the study of consciousness.

I will start with some general remarks about what motivates this approach, roughly what it is, and why it seems reasonable to think of it as both analytic and phenomenological. Then I’m going to argue for its legitimacy, by contrasting it with some other views with respect to what I see as a basic issue – namely, the use of first-person reflection in inquiry about consciousness. Specifically, I’m going to contrast my position on this with Daniel Dennett’s self-styled “hetero”-phenomenology and Eric Schwitzgebel’s skepticism about introspection. This will help me illustrate my approach, so as to further clarify and promote it.

To concentrate on methodology I will have to narrow the applications I consider. My focus here will be specifically on our approach to questions about how much richness or detail we ordinarily experience in vision. I choose this topic partly because it lends itself a bit better to brief discussion than others, and figures prominently in the two mentioned authors I will take as foils. This may seem to provide only an isolated, even toy example. But I think it is a significant issue, and I will eventually say a little about why.



Analytic phenomenology grows from a recognition that inadequate or divergent understandings of the terms we use to express questions about the mind hinder our ability to answer them. The obscurities and differences in interpretation of terms like “consciousness” or “representation”, for example, are not always minor or merely terminological, but sometimes reflect substantive disputes. And if sometimes they *are* largely terminological, finding that out can take some work, and play a key role in locating substantive issues. For these reasons, intellectual responsibility requires we try to explain carefully what we mean by such pivotal terms, as part of addressing the questions in which they appear. Analytic phenomenology takes up this challenge. But in clarifying distinctions, it also attends closely to concrete reality, and brings argument to bear on issues that give it a point. The issues I have in mind include: what consciousness is; how it is related to self-consciousness and intentionality; the character of perceptual, cognitive, emotional and agentive experience; and how these relate to mind, language, knowledge, selfhood, and value.

In this investigation, I make explicit appeal to first-person reflection. That is, in order to illustrate and thus to improve understanding of the concepts in need of clarification, I rely on first-person judgments applying them in both actual and hypothetical cases. Now, if such application of these concepts is really to improve understanding, it must enjoy some degree of warrant. And so I do hold that the first-person judgments relied on enjoy at least some degree of warrant, in the absence of contrary evidence. In other words, I operate under the assumption that we enjoy a default first-person authority for some of what we say about our own experience. But I do not, as a methodological matter, assume a particular positive view of what accounts for this authority, an explanation of what gives us a distinctively first-person warrant for judging about our experience. I believe such an account is possible, but can emerge only from the practice of phenomenology – so it is not where we start. However, one can, in advance of this, say something to legitimize my methodological assumption, and I offer to do so in what follows.

Please notice the assumption is *not* that some first-person judgments about experience are infallibly correct, or that they are incorrigible (in the sense of being warranted in a way that cannot be overridden). One or the other claim might be true, but it is not to be assumed at the outset. Though I see myself as continuing something of the phenomenological tradition of Brentano and Husserl, my initial methodological starting point are perhaps more modest than those of either the first (with his appeal to the infallibility of inner perception), or the second (with his appeal to apodictic evidence).

There is another potential difference worth noting between the kind of phenomenology I want to do and Husserl’s. On one reading at least, Husserl tried to

detach the epistemic standing of his philosophical claims from commitments to all assertions about particular spatial or temporal matters of fact, by universally suspending these in his famous methodological bracketing, or *epoche*. I do not want to enter into exegetical questions about whether this, together with a pure apriorism, is indeed fundamental to his phenomenology. I just will say it is not essential to mine. For I do not try to clarify the relevant distinctions in some way that floats free of our capacity to correctly apply them in concrete instances. And I do not think it desirable to insulate phenomenological inquiry from experimental and clinical findings in psychology.

But there is another part of the Husserlian tradition, as I understand it, I definitely do want to embrace. My primary interest lies not in making explicit concepts alleged to be *already at work* in pre-philosophical, ordinary or commonsense thought. Rather, the point is to *develop concepts that will best serve us* in addressing philosophical questions about the mind. These will of course grow out of an initial ordinary understanding and usage. But it is not my goal to provide an analysis of meanings in general circulation. I am not aiming to say what people generally (or what “we” – whoever “we” are) mean by words like “experience”, “know” and “perceive” much more precisely and explicitly than people usually try to do. My phenomenology is not in the business of that sort of “conceptual analysis”. However, my practice is “analytic” insofar as I do try to be explicit about what I mean and about the form of my arguments in ways I think would likely seem “analytic” to those familiar with the use of that label – and insofar as my arguments engage with philosophers who would likely attract this label, if only because they can be placed in a certain historical lineage. In saying this, I do not, however, mean to suggest that phenomenology in the Brentano-Husserl tradition is “purely descriptive” and so unconcerned with philosophical reasoning. Though again, I will leave exegetical questions to the side here.

In any case, phenomenological reflection as I conduct it is not purely descriptive in that sense, but involves an *open-ended critical process*. I do rely on the idea that some sincere first-person judgments about experience initially deserve respect. But I also see them as subject to critical reconsideration and revision, on the basis of further first-person reflection, in response to pointed questioning, and attention to relevant distinctions and implications. This critical process, incorporating a practice of first-person reflection, may either decrease or increase the initial default warrant presumptively granted specific instances of it. What my recommended practice of *first-person reflection* comes to will I think be a clearer if I now proceed, as promised, to contrast my stance with Dennett’s “hetero-phenomenology”.

On Dennett’s official method for studying consciousness, first-person reports and beliefs about experience are presumed to have *no authority at all* – unless

and until their accuracy is confirmed by third-person theorizing. This sometimes overlooked aspect of his view becomes particularly vivid in an analogy he draws with a fantasized anthropological study. Dennett asks us to imagine that the members of some cultural group worship a “tree god” they call “Feenoman”. Then he observes: good anthropologists studying the Feenoman culture would not grant any – even defeasible – authority to its members’ assertions about the existence and actions of their tree god. Analogously, if we are to be good consciousness researchers, we will grant no authority at all to any of a subject’s beliefs about his or her experience. Rather, while remaining thoroughly “neutral” or “agnostic” about their truth (Dennett 1991, p. 82–83), we should seek to determine what, if anything, in them is correct, by trying to explain them in the light of whatever evidence this neutrality permits us to gather from an observer’s, third-person point of view. And so, we are each of us to consider even our beliefs about our own experience as if they were another’s, to which we granted no authority, and of which we required external justification. In this sense all legitimacy for speaking of your own experiential life must derive ultimately from the viewpoint of another, whose interest in you lies entirely in explaining your observed behavior. Hence the “hetero” in “heterophenomenology.” Though Dennett acknowledges that to proceed in this way is alien to our practices in ordinary life, he believes that science demands it of us.

Part of my disagreement with Dennett lies in the fact that I do not think that the experimental study of experience – such as is found in work on imagery and visual illusions – actually does require his “no default first-person authority” epistemology. And we can reject his methodology without rejecting science. Right now, however, I will focus on a different, general concern. Dennett is confident that his method holds us to high standards without risk of losing anything valuable we should want. “Why *not* live by the heterophenomenological rules?,” he asks “[...] Nobody has yet pointed to any variety of data that are inaccessible to heterophenomenology.” (Dennett 2003, p. 39) I want to challenge this confidence. I have two objections.

The first one is this. According to heterophenomenology, I can (and should) study the beliefs I hold about my experience while remaining thoroughly neutral as to their truth. But this study of my beliefs can remain truly neutral only if I can understand what I mean by their expression well enough for the purposes of investigation without ever *relying* on their truth. Suppose, for example, I say I feel itchy. If Dennett’s methodology is correct, I should be able to understand *what I mean* by “itchy feeling” well enough to investigate and try to explain *my belief* that I feel an itch, without relying on my correctly telling when, if ever, I actually feel an itch. And indeed, the same would hold true for any and every notion with which I might purport to say how my experience is. To investigate *what I mean*,

and hence *what I believe*, in saying I feel pain, am troubled by persistent thoughts, amused by your remark, or see a Gestalt switch, I must depend not at all on an ability *to tell correctly when these happen to me*.

This assumption is highly questionable. For it is highly questionable whether normally you can do anything to demonstrate – to yourself or anyone else – how you understand, for example, “itchy feeling” in self-attribution, which doesn’t involve correctly identifying an itchy feeling in your own case. Would I understand “itchy feeling” as I do if, when I said, sincerely, “I feel itchy” I *never or rarely did feel itchy*? It is at least reasonable to think I would not. And I would maintain that the same holds, directly or indirectly, whenever we are dealing with concepts whose self-application suffices for a first-person judgment about experience. If this is right, it is reasonable to think you can *understand* experiential notions as you do in applying them to yourself only if you regularly apply some of them *accurately*. And of course you can know what beliefs you express only if you understand the terms you use to express them. So it is also reasonable to think you can know what beliefs you hold about your experience, only if you regularly form accurate ones. All this calls into question whether we can be strictly neutral about our capacity to form true beliefs about our own experience, and still maintain we understand what we *mean* in speaking of it, and know what we *believe* about it. So there is reason to suspect Dennett’s radical third-person approach will not, as he maintains, leave all the relevant data accessible. For, on reasonable assumptions, this approach, in refusing to grant we make true judgments about our own experience, renders experience unintelligible to us, and deprives us of the very beliefs it was supposed to study and explain. If one asks: “What variety of data does a heterophenomenological study of experience render inaccessible?”, the answer would be: *its entire subject matter*.

My criticism here is not that it’s fully clear that the relevant grasp of experiential concepts is in fact inseparable from their successful first-person application. Nor does analytic phenomenology assume that this is so. The argument is simply this. Suppose we have a choice between: (i) a methodology like Dennett’s that depends on a highly questionable assumption, which, if false, would render its ostensible subject matter unintelligible; and (ii) an approach like the one I propose that does not thus court its own self-defeat. Then, if there is no compensating advantage to (i), we should go with (ii). And there is no evident critical superiority to Dennett’s approach that would overcome this. For there is no legitimate criticism to which his approach subjects first-person judgment from which my alternative shields it. Thus we have reason to favor my method’s provisional reliance on subjective judgment over Dennett’s radical third-personalism.

Here is my second, related reason for this conclusion. Not only does Dennett’s approach risk alienating us from its own ostensible subject-matter, it

would take from us the very tools we need – by his own implicit admission – to think critically and responsibly about the issues it aims to address. To see this, consider what Dennett thinks naïve reflection tells people about their own visual experience. We tend to believe, he thinks, that “the visual field [is] uniformly detailed and focused from the center out to the boundaries” (1991, p. 53). Leave aside for now whether people actually do believe such a thing, or why Dennett thinks they do. Just suppose that someone does indeed hold such an erroneous introspective belief. Now ask: *how is one supposed to correct it?*

Dennett says the following “simple demonstration” will do the trick. First, fix your gaze dead ahead, while holding a previously unexamined playing card out at the periphery of your visual field. Then, without moving your eyes to look at the card, slowly bring it towards your point of visual fixation, and notice how close you have to get before you can make it out well enough visually to report what type of card – say a seven of diamonds – it is. But now notice: Dennett’s suggested procedure for correcting our alleged introspective belief itself *relies on introspection*. For the “simple demonstration” apparently presumes I am entitled to at least *some* of my introspective judgments about experience. If the demonstration is to work, I need to be warranted in saying things like “I am now looking straight ahead, not at the card” and “The card now looks nearer to the spot I’m looking at”. But I seem to have no right to rely on such judgments, if I am truly playing by “heterophenomenological rules”. For these rules demand total neutrality towards the truth of my first-person judgments about experience until they have been somehow vindicated from the outside in a way that yields nothing to first-person authority. Without such vindication, the most I have done is just to gather more beliefs about my experience, which have no standing to correct anything.

One might object: “Dennett isn’t using introspection to correct itself but to show it up as fatally liable to contradictions it can’t resolve”.<sup>1</sup> However, it is clear from the quote above that Dennett, at least, regards the demonstration as establishing the falsity of the initial claim of naïve reflection, and he seems to allude to this finding later when he remarks that he has earlier shown the richness of consciousness to be an illusion (1991, p. 362). You might, however, propose that we *could* use Dennett’s example to make introspection “self-destruct” (even if *he* does not). In response I would say that the “equally detailed all the way out” generalization about the visual field (if anyone really is tempted by it) comes from a failure to ask certain legitimate questions, which, once posed to reflection, readily disconfirm it. So the “always equal detail all the way out” and the “detail falls off a lot from the point of fixation” judgments – even if

---

1 Thanks to Chris Georgen, Eric Schwitzgebel, and Leopold Stubenberg for each pressing this point.

both are in some sense “introspective”, are not epistemically on par, since the second is based on considerations that were neglected in forming the first.

My point here is not to endorse Dennett’s “card trick” debunking of naïve reflection. In fact, as I will explain, I think it is flawed. Right now the point is this. The *sort* of thing Dennett is trying to do here is legitimate. He’s trying to show that some would-be introspective claim about the character of experience is mistaken, by appeal to *more searching* first-person reflection. He does this by focussing on implications of the target claim, and then reflectively interrogating experience in a previously neglected fashion, so as to correct a superficial initial notion with one more critical. This is fine by my lights – in fact, to do this is to engage in analytic phenomenology. The problem is that “hetero” phenomenology would prevent us from engaging in this legitimate practice, since it accords no epistemic status to the judgments the demonstration evokes that would give them the power to correct. Thus, if we really did follow the heterophenomenological rules – as not even their author actually seems to do –we would rob ourselves of the opportunity to improve introspection by critical reason. (And this, perversely, in the name of scientific rationality.)

I hope this clarifies to some extent the character of the analytic phenomenology I recommend, by contrasting it with Dennett’s radical third-personal methodology. The former is preferable to the latter for at least two reasons. First, what Dennett calls heterophenomenology threatens to undermine our understanding of the concepts of experience that figure in our first-person beliefs about it. So it threatens to deprive us of the subject matter it is supposed to reveal, and offers us no special advantages that would make this risk worthwhile. Second, a truly consistent heterophenomenology would paralyze the sort of critical first-person reflection by which (as Dennett himself recognizes) we rightly attempt to improve those beliefs. My approach, by contrast, suffers from neither problem: it does not invite self-defeat, and it affirms and encourages self-examination.

Let me turn now, as promised, from Dennett to a second sort of challenge to my approach. This lies in Eric Schwitzgebel’s withering assessment of introspection. He considers various questions about experience regarding which, he argues, introspection returns mistaken, uncertain or conflicting judgments. From this he concludes that introspection fails us: “[...] not only in assessing the causes of our mental states [...]. We make gross enduring mistakes about even the most basic features of our currently on-going conscious experience, even in favorable circumstances of careful reflection, with remarkable regularity.” (Schwitzgebel 2011, p. 118–119) Now this seems pretty discouraging. For apparently, even when conditions are *favorable*, and even when we strive to be *careful*, first-person reflection *persistently* and *regularly* leads to *huge* errors about even the most *elementary* matters. If Schwitzgebel is right, it would seem analyt-

ic phenomenology is doomed. For it tries to draw on an irremediably unreliable source of information.

I should say right away Schwitzgebel's full attitude here is more ambivalent than this suggests. Though he sees most as too timid in their critique of introspection, he also indicates he finds it crucial to the study of consciousness. (Schwitzgebel 2011, p. 118) And it is notable that he calls the introspection he disparages "naïve" – suggesting perhaps there may be another, *sophisticated* form that merits respect. However, Schwitzgebel does not explicitly embrace the idea that there is a respectable form of introspection, and tell us what it is. Rather, he seems to leave us with a kind of "can't live with it, can't live without it" perplexity. And if favorable conditions and carefulness won't help us, it is very hard to see how there could be a legitimate practice of first-person reflection.<sup>2</sup> So I want now to take up the challenge of showing how indeed there is, by examining one of Schwitzgebel's star witnesses against introspection: our reflective impression of the richness or detail in visual experience.

Somewhat like Dennett, Schwitzgebel finds just here a prime example of gross introspective blunder. However, I will argue that his treatment – also somewhat like Dennett's – ultimately points away from the negative conclusion that he draws from it, and sails us into the arms of analytic phenomenology.

Schwitzgebel poses the "richness" issue this way: how much, he asks, "presents" itself to you "at any one moment" as "clearly in visual experience" "with shapes, colors and textures all sharply defined"? (Schwitzgebel 2011, p.125) He thinks the answer that you are (introspectively) likely to deliver is this: a "fairly wide swath" of what's before you. Notice this claim isn't nearly as shockingly reckless as the "equal detail all the way out" one that Dennett tries to pin on us. Still, Schwitzgebel maintains what introspection tells us is far enough removed from the truth about experience to constitute a "gross" error about basics. The *truth* is that the *at-a-moment* area of clarity in visual experience is *not* "fairly wide". In fact, it is only "tiny" (Schwitzgebel 2011, p.126).

Now, just what is supposed to show people do hold the "wide swath" belief about visual-clarity-at-a-moment, and that it is grossly mistaken? Schwitzgebel's procedure is this. First, he instructs us, "fixate on a point [...], hold your eyes steady while you reflect on your visual experience" of what you're not visually "fixating" on. Then "take a book in your hands and let your eyes saccade around

---

<sup>2</sup> In correspondence, Schwitzgebel clarifies that he wants to hold the door open to a *kind* of carefulness that would make introspection legitimate, perhaps involving the sort of training that Titchener discussed in his manuals from the early 20<sup>th</sup> century. Still, Schwitzgebel seems ultimately non-committal. At any rate, I wish to explain here how the *right kind of care* has more to do with philosophical sensitivity than with laboratory technique.



its cover while you think about your visual experience in the regions away from the precise point of” fixation. At this point, Schwitzgebel says most people seem to discover “that [...] the center of clarity [of visual experience] is tiny, shifting rapidly around a rather indistinct background”. And he says, they “confess to error in having originally thought otherwise” (Schwitzgebel 2011, p. 126).

I myself doubt that Schwitzgebel really has shown that people ordinarily go around forming grossly erroneous beliefs about how much they clearly experience visually in an instant.<sup>3</sup> But first I want to make a point parallel to the one I just emphasized about what Dennett is trying to do. I agree with Schwitzgebel that we should not trust just whatever claim about our experience pops into our heads on first consideration of just whatever question is thrown at us. We should instead consider the possibility of error in such judgments and seek to confirm or correct them, by posing the right questions and reconsidering the matter. Whether or not Schwitzgebel has found just the right questions and drawn the right conclusions, I applaud what he is *aiming* to do here. But notice now: this seems ultimately to undermine his official thesis. If in fact we can correct the “wide swath in a moment” claim by the means Schwitzgebel proposes, then we should not, after all, say with him that introspective error endures even in “favorable circumstances of careful reflection”. For whether the circumstances are favorable and the reflection is careful depends on *what questions we pose and how well we understand them*. The alleged error arises in *unfavorable* circumstances that are *deficient in careful reflection*, and is *dispelled* when we introspect *carefully* in more *favorable* circumstances. Schwitzgebel then is, in effect, proposing a *good* introspection to correct errors to which the *careless* sort gives rise.

So, it is not that introspection is some hopelessly error-prone faculty that cannot right its own wrongs even if we try to use it with care. Rather: while we may be disposed to make inaccurate judgments about our own experience, there are ways to engage in first-person reflection carefully, so as to either replace an initial erroneous pronouncement with something more accurate, or at least avoid claims we have no business making.

As I’ve said, although I gladly embrace the general idea of self-correcting first-person reflection, I’m not happy with how Dennett and Schwitzgebel try to implement it. I now would like to say a little about this. This will allow me to further illustrate the practice of analytic phenomenology as I understand it.

First, a word about Dennett. He thinks our surprise at the weakness of our peripheral vision in his playing card demonstration shows that we usually think we

---

<sup>3</sup> In recent correspondence, Schwitzgebel tells me that his considered view is that people rarely make such judgments.



see a lot more detail than we do. But it doesn't really show this. Much less does it show that we think that everything that appears to us from the center to the boundaries normally appears to us in equal detail and focus. We do think – correctly – that normally we can tell quite a bit about shapes, sizes, and colors visually. But we often don't realize how much this depends on our restless, naturally wandering gaze. For we *often take very little notice* of how much we ordinarily shift our gaze to see as much and as well as we do. And so, when artificially, we deliberately suppress our near-tireless impulse for visual exploration, and freeze our gaze, we may well expect that the ease with which we usually visually identify things will carry on pretty well. But that expectation does not hold up. This explains our surprise better than this “equal focus and detail all the way out” notion Dennett imagines we hold. I conclude his purported correction of first-person reflection fails, for failure to find an appropriate target.

Now to Schwitzgebel's argument. First, I would be skeptical about his idea that people, without some prompting, really do regularly form judgments about the size of the area of visual clarity they experience in little temporal snippets. I doubt we even can, on the basis of introspection, make sufficiently consistent judgments about the “area of clarity” size experienced in uniform split second intervals to accurately compare these. What I *don't* doubt is that Schwitzgebel's subjects, when they followed his instructions, began to notice something they *didn't* notice before about their visual experience. They begin to notice their gaze is shifting around on what appears before them much more quickly and often than they noticed before. And they begin to notice, more than they did previously, that there are frequent, rapid changes in how well or clearly they see the shapes and spatial arrangements of what appears before them. But I don't see why they should interpret this as forcing a big abandonment of previously held beliefs about how much they saw when, rather than a discovery, through heightened reflection, of facts previously ignored.

Let me take stock. Dennett's total subordination of first-person judgment about experience to a third-person perspective, and Schwitzgebel's views about the hopeless unreliability of introspection, would apparently rule out the practice of analytic phenomenology. For this relies on a default, but correctible first-person authority in the investigation of consciousness. I have defended analytic phenomenology against these authors' views. In the process I have shown how they implicitly partake of it, in evident defiance of their official positions. And while I approve of their phenomenological impulses, I have expressed misgivings about just how they express them. That is, I criticized the way they ask us to interrogate our experience, in order to form an accurate reflective understanding of its character – an understanding specifically, of how quickly how much of our environment becomes visually apparent to us. And I criti-

cized the inferences they drew from this interrogation to impugn “naïve” introspective judgment.

Now suppose I am right about all this. Still, what is my *positive* idea about how to examine visual experience to address questions about its richness? And what results do I hope to get out of this – if not the usual belittlement of introspection?

First, here is how I would recommend reflecting on your visual experience to think about these “how much how fast” questions raised in Dennett and Schwitzbeigel. Let’s begin by bringing this talk of “visual fields”, “saccading” and “areas of clarity” of “visual experience” a bit more down to earth, by approaching the matter with a little ordinary word, “look”. Let’s agree that one way to report your “visual experience” of something, as distinct from your belief or judgment about it, is to talk about its *looking* some way to you – and when it looks somehow to you, it appears somehow to you. Now notice in concrete cases that you often change what you’re looking *at* and where you’re looking. You do not often look at one thing fixedly, without looking at diverse parts of it for very long. The gaze is typically restless, moving on – to this, now this, now this. And, as your gaze shifts, as you look at different things, as where you look changes, the way things look to you also changes. As this happens, not all of what’s before you that looks to you somehow is something you are also simultaneously looking *at*. For example, when you’re reading, the words on the page you’re looking at are not all that then looks somehow to you. Some of the area ahead and behind and around them also looks somehow to you, as your gaze capers along. Now, arrest your gaze on a word, and ask yourself, does it not look or appear differently to you now than any word (even the same word) appears to you when it is to the side of what you’re looking at? More generally, is it not the case that what you’re looking *at* appears quite differently to you than what you’re *not* looking at, but lies in an area that still looks somehow to you? Specifically, would you not say that what you’re looking at appears to you *in much more detail, more clearly*, its shape, boundaries, location are *more apparent* to you, than what you are not looking at?

Let us assume that your answer is yes. Return now to the notion that Dennett attributes to us – that the visual field appears in equal detail and focus all the way out from center to the boundaries. What you’re looking at is at the “center of your visual field”, I presume. And the area before you that you’re not looking at, but which is still *apparent* to you, includes the “out to the boundaries” part of the “field”. But then the idea that Dennett thinks we find introspectively appealing should, on the contrary, strike us as introspectively *preposterous*. For it should be clear to us from first-person reflection that how much is apparent to us of the spatial

features – the shape, location, size – of what looks somehow to us over time varies *enormously* with whether we are, at any given time, looking at it.

Suppose you arrive at some such phenomenological reflections. What lesson do I wish to suggest you draw? First, I think they show how responsible, critical first-person reflection, in reasoned dialogue with others, is feasible, after all. Second, even if (as I claim) they don't reveal and disabuse you of some huge illusion about yourself, still, this doesn't mean they leave you just where you started. For they may awaken a heightened appreciation of the character of your visual experience, which prepares you to consider issues that might have escaped you before. *That* is the proximal result that phenomenology aims to achieve – a new or enhanced reflective awareness of how you experience things that primes you for further thoughts – thoughts that promise to place what you now notice in the context of reasoning that gives it positive philosophical import, and draws you on to yet further such noticings.

All right, but just *what* further thoughts might these starting reflections facilitate, on my view? Their positive import lies largely, I think, in their value as a starting point for thinking about the intentionality of visual experience – that is, its directedness or reference to an object. I think this because I think once you have noticed the kind of change in appearance I've been discussing, you are ready to go on to notice the following. As you change what you're looking at, how it looks to you changes – how, for example, its shape looks to you varies. But what appears to you does not then appear to you to *change shape*, but to remain the same in shape. In other words, there is an experience of shape constancy in vision through attentional variation in perspective. This phenomenal constancy, I would argue, provides the phenomenal basis for the intentionality or object-directedness of experience: this makes it possible for visual experience to make apparent to us objects that “hold fast” amid changes in our experience, and thus that go beyond or transcend its perspectival limitations.

So a phenomenological response to the “how much experience” issue prepares us for an understanding of *constancy*, which illuminates the *objective reference* (or intentionality) of visual experience. Once we have this, we can begin to address questions about its *epistemic role* – how sense experience *warrants judgment*, and how it enables us to *understand demonstrative reference*. And we can investigate not only how experience yields objects for us to judge but how it enables us to confirm or correct our *classification* of them – we can explore the phenomenology of perceptual *recognition*. What we find here can then recast our approach to how we think and know about experience itself. To what extent, if at all, does the phenomenology of self-awareness parallel that of sensory awareness of one's surroundings? Such inquiry bears further on just what experience – consciousness – is. For

example, it bears on issues regarding what sort of self-consciousness, if any, is inherent to any consciousness whatsoever.

This indicates something of the path my own thought has been following in analytic phenomenology (in, e.g., Siewert 2012, 2013, 2015). I explicitly admit that I don't pretend to be breaking new trails here. Many of the points I have been making – or hinting at – have recognizable analogues in Husserl, and maybe should even count as reformulations of things he said. But it is not alien to analytic phenomenology to explicitly seek to appropriate, revise and reformulate its intellectual heritage. I would just say we need to take care that textual exegesis does not *substitute* for first-person reflection. Otherwise we are not ourselves *doing* phenomenology, but only commenting on phenomenologists.

I hope to have done something to explain, defend and illustrate analytic phenomenology as a philosophical approach. In closing, I want to comment briefly on something that often comes up when the use of introspection is considered, and that commonly worries philosophy as well – the elusiveness of consensus.

It is true that there are general questions about experience about which first-person reflection does not readily produce agreement, even where it seems implausible that this simply reflects individual variation in experience. And Schwitzgebel is right to note that recent discussion of so-called cognitive phenomenology provides a case in point. But that is not because of some deep defect in “introspection” that prevents our using it responsibly. Rather it has to do with whatever it is about *philosophy* that seems to promote interminable disputation, and makes it difficult. For this reason, contrary to what Schwitzgebel suggests (Schwitzgebel 2011, p. 128), we shouldn't expect introspection always to be “easy”, wherever it genuinely has a place. For some questions, the right conduct of first-person reflection (“introspection”) involves philosophical skill. So the sources of the relevant disagreement lie, I think, in our inadequate or divergent pursuit of philosophical inquiry. This is evident in the case on which I've focused – the differences in the views Dennett, Schwitzgebel and I take on the question about the character of visual experience. We all rely in some way, it turns out, on first-person reflection in constructing our views. These diverge, not because we each have in our heads some introspector mechanism that weirdly churns out arbitrarily diverse and incompatible answers to the very same questions about the mental life we all share. They diverge rather, because we ask different questions, proceed from different background assumptions, and operate with different or differently interpreted terms – influences to which we can all be oblivious, even willfully blind, and stubbornly committed.

This style of analysis can be extended to other cases (about emotion and thought) Schwitzgebel uses to illustrate introspective failure. He asks, for example: whether emotions occur unfelt; whether they are always experienced to have

bodily location; whether they consist entirely in feelings of bodily arousal; and whether there is a distinctive phenomenology of thought. In each case he invites us to note introspective uncertainty and disagreement. While a full discussion of these cases is not possible here, the form of a response is clear. We first should see that to answer Schwitzgebel's queries through reflection, we need to address *further* questions, like these: is it even appropriate to expect introspection to detect *unfelt* emotions? What do we include under the term 'feeling,' and what do we count as "emotion"? What would it mean for emotional experience just to "consist in" states of bodily arousal? What's the relation between this issue and the question about "localization"? Do these issues depend on how one resolves the question about "cognitive phenomenology"? And just what is *that* issue – what is implied by saying that thought has a phenomenology of its own, or that, on the contrary, the phenomenal character of thought is reducible to that of purely sensory states?

We can account for the disagreements or uncertainties Schwitzgebel's questions occasion by noting that these can be answered properly only by addressing the additional ones I just listed, and these – like a lot of philosophical questions – are hard to answer, and easily neglected. Once their relevance is acknowledged, then patience, serious intellectual effort, and at least a little creativity must be applied to answer them. They are, in any case, the sort of questions that will be approached differently by different people, who may reasonably disagree about how well each has done the job (provided they even take serious notice of each other's efforts). Thus if Schwitzgebel's questions occasion doubt and dispute, this is due not to some special flaw in "introspection", but to the fact that answering them requires we engage in philosophical – what I have elsewhere (Siewert 2011) called "Socratic" – introspection. And that is something we are liable to do inadequately, unequally, and divergently.

It may thus be unrealistic to hope that the relevant differences can be sufficiently exposed and ironed out to secure broad and deep professional consensus about all or most basic matters regarding consciousness. If that is so, then neither phenomenology, nor philosophy generally, will ever be anything like a textbook science, but will likely remain rife with idiosyncratic clashes and misunderstandings, and plagued by failures of attention and imagination. Even so, at least we can reasonably strive individually for some success in overcoming these problems, mustering as much integrity as we can. And we ought to remember that – particularly in philosophy – disagreement can also be a sign of intellectual vitality.<sup>4</sup>

---

<sup>4</sup> I am grateful for the interesting feedback I received on this material from the audience for my

## Bibliography

- Dennett, Daniel (1991): *Consciousness Explained*. Little, Brown: Boston, MA.
- Dennett, Daniel (2003): “Who’s on First? Heterophenomenology Explained”. In: *Journal of Consciousness Studies* 10, pp. 9–10.
- Schwitzgebel, Eric (2011): “The Unreliability of Naïve Introspection”. In: *Perplexities of Consciousness*, MIT Press: Cambridge, MA.
- Siewert, Charles (2011): “Socratic Introspection and the Abundance of Experience”. In: *Journal of Consciousness Studies: Describing Inner Experience: a Symposium Debating Descriptive Experience Sampling (DES)* Vol. 18, No. 1, pp. 63–91.
- Siewert, Charles (2012): “On the Phenomenology of Introspection”. In: Declan Smithies/Daniel Stoljar (eds.): *Introspection and Consciousness*. Oxford: Oxford University Press, 2012,.
- Siewert, Charles (2013): “Phenomenality and Self-Consciousness”. In: Uriah Kriegel (ed.): *Phenomenal Intentionality*. Oxford: Oxford University Press, pp. 129–165.
- Siewert, Charles (2015): “On Getting a Good Look: Normativity and Visual Experience”. In: Thiemo Breyer/Maxime Doyon (eds.): *Normativity in Perception*, Basingstoke: Palgrave Press.

---

presentation at the 37<sup>th</sup> Wittgenstein Symposium in Kirchberg, Austria, and from my students at Rice University. Also, special thanks to Steve Crowell, Chris Georgen, Sonja Rinofner-Kreidl, and Eric Schwitzgebel for their helpful comments.



Steve Fuller

# Towards a New Foundationalist Turn in Philosophy: Transcending the Analytic-Continental Divide

**Abstract:** I present a defence of foundationalism as a signature attitude in philosophy. I begin with an account of my own way into the topic, which was to do with trying to reconcile the two leading approaches to foundationalism in the post-Second World War period, namely, analytic and continental philosophies. What makes both ‘foundationalist’ is their requirement of a ‘clearing’ for thought prior to philosophizing. In practice, this means a removal of all *prima facie* epistemic privileges from the special sciences, a position that had characterised such ‘school philosophies’ as Neo-Kantianism and Neo-Thomism. The failed foundationalism of early 19<sup>th</sup> century German idealism plays a pivotal role in this narrative, with its path-dependent sense of teleology, in which alternative possibilities are eliminated over time. (This was the original meaning of ‘epistemology’.) This restricted sense of teleology proved to be modern foundationalism’s Achilles’ heel, not least because the 20<sup>th</sup> century world-historic events cast doubt on the feasibility – and even the desirability – of a world-order unified under the scientific method; hence, the varieties of ‘postmodern’ anti-foundationalism expressed in the final quarter of the 20<sup>th</sup> century. However, in the final section of the paper, I use the two main modern approaches to the empirical study of human beings – hermeneutics and naturalism – as a launch pad for renewing the foundationalist project with a much more expanded and open-ended (‘reversible’) sense of agency and teleology, one fitting for the being I’ve called ‘Humanity 2.0’.

## 1 My route into Foundationalism and Idealism’s pivotal role

Throughout my professional career it has been common to both recognize and decry the analytic-continental divide in philosophy. As someone who was trained in the analytic tradition (at Columbia, Cambridge and Pittsburgh) but whose philosophical interests more closely matched continental aspirations, I spent much of my time as a Ph.D. student trying to bridge the divide – that is, when I wasn’t working on my Ph.D. thesis (Fuller 1985)! My general strategy



was to recast the continental side in analytic terms, which seemed sensible, given the balance of power between the traditions in the American academy. For example, one of my first pieces, jauntily entitled ‘A French science with English subtitles’, tried to make Derridean deconstruction sound like Michael Dummett’s brand of antirealism (Fuller 1983). In this venture I was inspired by Richard Rorty’s more publicized manoeuvres along related lines. Actually, Rorty served as a negative example for me, in that I thought that he sold the continentals short by reading them as simply commenting on the history of philosophy, which in an anti-Quinean gesture he happily equated with philosophy as such. To be sure, Rorty was probably projecting his own professional ascent, since he was hired at Princeton to teach the history of modern philosophy, which amounted to slumming it in the top ranked – and resolutely analytic – American department (Gross 2008). This helps to explain the structure of Rorty (1979), which began life as notes prepared for his history classes.

For my own part, philosophers who were coming into their own in the early 1980s, such as Jon Elster (1979, 1984) and Ian Hacking (1975a, 1975b), appeared to be much better bridge-builders in their willingness to insinuate continental concerns more directly into analytic ones. They were not simply trying to establish the convertibility of continental and analytic positions, which was Rorty’s *modus operandi*. Rather, they were showing how continental insights could nuance analytic positions in some presumed common philosophical project. At least, that is how I understood matters thirty years ago. (Perhaps I was too optimistic?) In any case, the existence of a common philosophical project between the two traditions perhaps explained their mutual antagonism, an animus that did not extend to, say, Neo-Kantianism, Neo-Thomism or, for that matter, Pragmatism – movements that were tolerated on all sides in the sense of being politely ignored. What the self-identified ‘analytic’ and ‘continental’ philosophers share is a vision of philosophy as concerned with – for a lack of better word – *foundations*. Even when philosophers in these two traditions grant that we must begin philosophizing *in medias res*, we should somehow try to access some ultimate level of being. The path to this state may be paved by heroic abstraction, linguistic reflection or logical presupposition. But once that point is reached, then philosophy properly begins.

One telling mark of foundationalism as a philosophical attitude is to note the reluctance of Bertrand Russell, despite his avowed empiricism and naturalism, to embrace William James’ observation that the mind’s default state is one of ‘blooming, buzzing confusion’ (Russell 1946, chap. 29). It was not that Russell failed to recognize the experience. Rather, he refused to accord it philosophical significance. For Russell, this state of confusion was not itself a starting point but a problem to be resolved in some fashion, so that philosophy may then begin in

earnest. It is this gesture – that the mind’s clutter needs to be cleared before philosophy takes flight – that is the mark of the foundationalist turn. In this respect, the early 17<sup>th</sup> century founders of the scientific method, Bacon and Descartes, were clearly foundationalists by virtue of their desire to strip away the prejudices and traditions in which people were normally immersed and which inhibited the expression of their God-given powers of reason. By the early 18<sup>th</sup> century, Locke had articulated a proto-positivist version of foundationalism when he described himself as an ‘underlabourer’ who clears the way for the true founders – the ‘master builders’ of thought, whose job description better fit his friend Newton’s than his own. My long-standing disagreement with Locke on this point lies simply in the self-deprecating role he assigns to himself in a vision of knowledge that I otherwise largely accept. Rather like Popper, I hold that the division of labour between the philosopher and the scientist should not be so sharp – the latter is simply a technologically enhanced version of the former (Fuller 2000: chap. 6.)

Thus, the most striking feature of foundationalism is its requirement of a *clearing* before philosophy can happen. In other words, it presupposes a sort of space in which an intellectual edifice (aka knowledge) can be built. In the late 17<sup>th</sup> century, this space was formally recognized by the Cambridge Platonist Ralph Cudworth as ‘consciousness’. The term was designed to capture the character of the place where humanity interfaces with the cosmos. In the Christian context, this meant our relationship with God. There are two general ways of thinking about this space, both of which can be found on either side of the analytic-continental divide. In one case, consciousness is the stage in which you are the dramatist; in the other, you are the spectator of what transpires on the stage. In the former, we live up to our divine entitlement by creating in consciousness as God creates in reality more generally, which is to say, as a form of self-affirmation. (Putting it this way nowadays conjures images of ‘virtual reality’.) In this context, consciousness is the site in which human agency is enacted, where decisions of existential import are taken. Here Descartes and Pascal – often portrayed as antagonists – find themselves on the same side: Where Descartes posits the *cogito*, Pascal makes a leap of faith. The other way is to see consciousness as a site for the delivery of being or the revelation of truth – *aletheia* in Heidegger’s sense, but also ‘evidence’ in the sense invoked by analytic epistemologists to mean a self-certifying experience (drawn from Brentano’s *Evidenz*). In this case, the human is more passive, submissive, even abject. Nevertheless, this vision is arguably more in line with the receptiveness required for ‘enlightenment’ in the great Eastern religions and even Islam, in which Muhammad is presented principally as a divine vessel.

While the 17<sup>th</sup> century thinkers were largely responsible for a conception of the mind that permitted this foundationalist turn in philosophy, it is Kant to whom we owe the modern idea that philosophy provides the foundation for all the other academic disciplines. This was a role previously occupied by that ‘queen of the sciences’, theology, which was increasingly ill-suited to a rapidly secularizing world, in which the classic professions of law and medicine were vying for academic superiority. Kant’s polemical contribution to this discussion, *The Contest of the Faculties* of 1798, inspired Wilhelm von Humboldt to re-launch the University of Berlin as a vanguard national university with philosophy enshrined as the unifying subject for all courses of study. This was the environment in which German idealism flourished, with Fichte, Schelling and Hegel informing the paths that both the human and the natural sciences took in the second half of the 19<sup>th</sup> century, the period when modern academic disciplines began to acquire their present shape. Indeed, this was probably the period when philosophers had the most substantive impact on the development of first-order fields of inquiry (Fuller 2013).

German idealism is rarely given credit for having seeded the panoply of modern disciplines, notwithstanding the valiant effort of Cassirer (1950). On the contrary, the movement has been stereotyped as scholastic and even anti-scientific, mainly because its seemingly rigid adherence to the ‘dialectical method’ placed *a priori* limits on the path of disciplinary development – specifically by declaring the ‘impossibility’ of various properties and objects that were later demonstrated to exist. The most notorious cases in point concerned mathematics, including transcendental numbers, infinite sets and, perhaps most strikingly, non-Euclidean geometry, which paved the way for Einstein’s revolution in physics. It is hardly surprising, then, that originators of the most recent round of philosophical foundationalism – the ‘analytic’ Gottlob Frege and the ‘continental’ Edmund Husserl – both trained as mathematicians in the late 19<sup>th</sup> century, a period when the field sought independence from practical concerns in order to explore the conceptual limits of its own objects (Collins 1998, chap. 13). To be sure, Frege and Husserl charted the appropriate ‘transcendental’ standpoint rather differently: Whereas Frege followed the conditions under which a judgement is made in possibility space (i.e. logic as algebraic function), Husserl abstracted from an actual object to capture its essence, understood as all the conditions under which an object can appear: the former top-down, the latter bottom-up. They form of the two halves of the Cartesian legacy, which together presented philosophy as canvassing a universe well beyond the delivery of experience and its default pre-suppositions (cf. Solomon 1977).

Nevertheless, there is no denying that the original German idealists tied their philosophical foundationalism to an image of the ends of knowledge becoming

more conceptually specified, integrated and unified over time – not one open to new and potentially strange possibilities. Indeed, the idealists operated with a conception of progress that presumed path-dependency, a focusing of world-historic options that was equated with reality's ever greater self-realization. Where the idealists differed – quite substantially – was over the correct characterisation of the agent of this process: Ego (Fichte), Nature (Schelling) or Spirit (Hegel). Ironically, the one legacy of idealism that survived intact in the 19<sup>th</sup> century was the functional differentiation of disciplines – except without the overarching systemic vision that was supposed to be philosophy's unique epistemic contribution (Schnädelbach 1984, Chap. 1). Indeed, James Ferrier coined 'epistemology' in English as the leading branch of metaphysics in the mid-19<sup>th</sup> century to capture this point just as it was losing its institutional force (Fuller 2015, Introduction).

Without philosophy's second-order comprehension of interdisciplinary divisions, which had been idealism's signature style of foundationalism, a more complex version of the situation that Kant addressed in *The Contest of the Faculties* was allowed to re-emerge in the second half of the 19<sup>th</sup> century. These conflicts were characterized as *Werturteilsstreit* and *Methodenstreit* in the German-speaking world, especially when the battles were fought at the interface of the human and natural sciences (Proctor 1991, Part II). Typically the dynamic of these disputes resembled jurisdictional conflicts, in which attempts were made to extend the proper empirical domain of disciplines into totalizing world-views – as if to fill the philosophical void left by idealism. (If in doubt, read the multi-volume Merz 1965.) In the 20<sup>th</sup> century, terms such as 'reductionism', 'scientism' and even 'imperialism' came to be used to capture this spirit, usually with quite specific disciplines in mind, especially physics and economics. However, the late 19<sup>th</sup> century witnessed much more of a 'war of all against all' attitude, in which, say, 'psychologism' functioned as a multi-purpose shibboleth to impugn virtually any discipline's or school's scientific credentials, or claims to objectivity more generally (Kusch 1995). From this standpoint, the cross-disciplinary allegations of 'relativism' raised in the late 20<sup>th</sup> century might be seen as an ironic reversal of this strategy, whereby disciplines or schools come to be accused of not being sufficiently imperialistic in their epistemic ambitions!

In the context of a downsized, post-idealist philosophy, the most that philosophers can hope to do is to referee cross-disciplinary disputes by aligning each discipline with a set of distinctive methods and objects which, when taken together, enables a peaceful coexistence in the academy without requiring idealism's superordinate sense of epistemic authority. Neo-Kantianism and Neo-Thomism were the secular and sacred versions of this orientation that developed in the late 19<sup>th</sup> and early 20<sup>th</sup> centuries. Both are very much 'school philosophies'

for whom epistemology recapitulates academic bureaucracy (Fuller 2009, chap. 2). However, the institutional dominance of non-foundationalist philosophies was effectively subverted in the First World War, when Germany, the nation with the most advanced higher education system, suffered a devastating defeat, despite its widespread support across the academic community. This included not only the human and natural sciences but also theology, which had played a formative role – via the higher education minister and prominent Church historian Adolf von Harnack – in establishing the original state-industry-university (‘triple helix’) complex, now known as the Max Planck Institutes. It was this perhaps justified – but ultimately defeated – confidence that sparked a new era of foundationalism in the postwar era, most notably logical positivism and existential phenomenology. Common to these two movements – notwithstanding their increasing differences following their Nazi-provoked transatlantic exile – was greater attention paid to the ‘sceptic’ as a philosophical opponent. Thus, early modern figures such as Descartes, Hume and Kant were given a new spin in the philosophical canon: less about paving the way to science and more about assuaging the sceptic’s doubts.

## 2 Reappraising and updating Idealism’s legacy

In most histories of philosophy the need for an intellectual clearing prior to the laying down of epistemic foundations is presented as going against the grain of the idealists’ brand of philosophizing. However, on closer inspection, this difference is of a rather specific nature, since both are indeed foundationalists. Russell followed the early modern line – championed by Bacon, Descartes and Locke – that treated the clearing of the mind and the provision of foundations as two distinct and separate tasks. Nowadays we would assign the first task to a kind of cognitive psychology-led pedagogy, which aims to divest us of our biases and liabilities, and the second task to the rigorous application of ‘the scientific method’, however that is defined. Unsurprisingly, ‘philosophy’ as a discipline has a minimal role in this enterprise, other than ushering people from one phase to the next. In contrast, what Russell saw as two tasks, the Idealists saw as proceeding together as two indivisible parts of one overarching task, an ongoing disentanglement and purification of one’s understanding of the world. Equally unsurprisingly, ‘philosophy’ now turns out to be the discipline that encompasses all disciplinary knowledge. Yet, both Russell and the idealists would be in agreement that philosophy’s endgame should involve the most comprehensive, systematic, rational vision of the world, in which science is brought to ul-

timate fruition. Considering the various assaults on the supremacy of ‘science’ in our postmodern times, this is not a trivial point of convergence.

To appreciate idealism’s *modus operandi*, one may think at the dawn of inquiry that unicorns exist because we can form an idea of a horned horse because we’ve seen horses and we’ve seen horned animals. Such imaginative associations are the stuff of myth and fiction. However, as we learn more biology, we realize that such a thing could not exist, given the actual course of evolutionary history, etc. In other words, as we weave together our expanding empirical base under a common conceptual framework, what we originally thought were contingent facts (e.g. it so happens that unicorns do not exist, but they could have existed) metamorphose into something more closely resembling rational necessity (i.e. evolution makes it highly unlikely that unicorns would come into being). Indeed, Spinoza’s slogan, adapted from the Stoics, is apposite: ‘Freedom is the recognition of necessity’. Human progress consists in learning to desire that which we will come to discover must be the case, so that by the time we come to the ‘end of history’ there is no discrepancy between expectation and outcome, as belief converges with fact in absolute knowledge.

This account of the idealist *modus operandi*, clearest in Hegel, was revived by Alexandre Kojève and Alexandre Koyré in 1930s Paris, and reinvented by Francis Fukuyama (1992) for post-Cold War America in the 1990s. In both these 20<sup>th</sup> century incarnations, the idealist project was presented as a foundational course of study for aspirant philosopher-kings or, as the novelist Tom Wolfe sarcastically called them in the 1980s, with a tilt to Wall Street, ‘masters of the universe’. Indeed, by the time of Fukuyama, not only had capitalism triumphed but also Communism had come to be seen as, so to speak, a political unicorn with no chance of sustainability. Against the backdrop of this line of thought, so-called ‘post-structuralist’ or ‘postmodern’ strains in recent French thought, which have done so much to shape ‘critical theory’ in the English-speaking world for the past half-century, should be understood as an internally diverse backlash (cf. Descombes 1980). What Kojève and Koyré, two émigré Russian Jews, promised in the 1930s was an alternative reading of world history that delivered a different outcome from the one that the Nazis predicted. In one obvious sense, they were proved correct. However, the character of the Second World War’s ending – mass aerial bombings of Germany and Japan, culminating in the dropping of two atomic bombs – immediately caused a crisis of confidence on the idealist side, which throughout the Cold War was given most articulate expression in Jean-Paul Sartre’s brand of Existentialism. The ‘postmodern turn’ involves suspending a key assumption of this entire trajectory, namely, that history has an ‘end’ that somehow resolves diverse claims to legitimacy in a single ultimate regime with absolute authority – what epistemologists call ‘The Truth’.

Nevertheless, what makes ‘postmodernism’ in this broad sense interesting to ‘foundationalist’ thinking in philosophy – and why it deserves the unqualified honorific term ‘theory’ for its activities in the English-speaking humanities – is that it continues to operate at a ‘big picture’ level that tries to probe more ‘deeply’ into matters than ordinary empirical inquiry would normally permit. If we take the two strands of modern philosophy that Kant identified at the end of *The Critique of Pure Reason*, the idealists may be seen as having stressed the *rationalist* side and the postmodernists the *empiricist* side of the divide. Moreover, Kant’s original discussion (where the ‘empiricists’ are represented by Epicurus) gets to what is at stake in the contemporary discussion. The rationalists are committed to a foundation based on rules (or ‘laws’) that can be laid down by some means or other. While we may not have created these rules (God perhaps did), we nevertheless are in a position to enforce them, once they have been discovered. On the other hand, the empiricists believe that the only rule that we shall ever discover is that there are always exceptions, in turn reflecting the ultimate elusiveness of reality from our grasp and control. Thus, where rationalists veer towards dogmatism, empiricists veer towards scepticism in their respective accounts of foundations.

A brief look at five major strains of broadly ‘postmodern’ thought in recent France brings this point out well. All are ‘empiricist’ in the sense that Kant meant:

1. *Derrida* – Modernity self-destructs conceptually because it can only repress but never fully contain whatever opposes its dominant tendency.
2. *Foucault* – Modernity self-destructs empirically because it becomes unsustainable in the face of overwhelming anomalies to its regime of knowledge-power.
3. *Deleuze* – Modernity mutates into something unexpected, perhaps even perverse, yet it remains an outgrowth of modernity’s original logic.
4. *Lyotard* – Modernity is falsified by the history that it claims on its own behalf, which does not have the teleological character that modernists claim.
5. *Latour* – Modernity is an organized form of self-deception by which humans try (unsuccessfully) to separate themselves out from the rest of nature.

But suppose one remains sympathetic to the idealist take on foundationalism – as I do. Can we make a less totalizing diagnosis of its problems than our five French postmodernists do? Two interrelated but potentially tractable problems come to mind.

The first problem is the one that 19<sup>th</sup> century thinkers had already understood, which is related to idealism’s *apriorism*. Apriorism has been historically tied to infallibilism (aka epistemic certainty), perhaps in deference to the deity whose path we recreate when we think about things from first principles. Yet,



logically speaking, it is not self-contradictory to say that inquiry must begin with clearly stated universal propositions, which nevertheless may need to be revised, perhaps even radically, as they encounter the particulars of empirical reality. (Indeed, from a strictly theological standpoint, this may just be the right pose to strike, given Original Sin.) Of course, this is just another way of characterizing what Popper called the ‘hypothetico-deductive method’, precedents for which can be found in William Whewell’s version of Kant. Thus, science operates by a kind of fallibilistic apriorism, which Charles Sanders Peirce famously characterized as ‘corrigibilism’ (Laudan 1981, chap. 14). Yet, like the idealists, Peirce continued to believe that all paths of inquiry, however divergent in their hypotheses, would ultimately converge on the same representation of reality, which he operationalized as the consensus of scientific opinion. Even Popper seemed to back a version of this view, albeit in a more muted and vaguer way – and perhaps in spite of his vaunted anti-historicism (Popper 1957). In this respect, Kuhn (1970) posed an interesting challenge to this entire line of thought by introducing the Darwin-inspired idea that the history of science – if not history as such – may indeed be path-dependent, and in that sense ‘irreversible’, yet without heading to a specific predetermined end-state. Instead, the same original intellectual trajectory may be subject to refraction as it encounters various local obstacles to its passage, resulting in an increasingly complex and disunified scientific enterprise. So on what terms, if any, is a unification of knowledge possible (cf. Fuller 2007, Part II)?

This brings us to the second problem, which is related to the idealists’ over-estimation of history’s irreversibility. It led them to read the arrow of time as a relatively straightforward empirical indicator of history’s ultimate end. Thus, dominant tendencies were widely seen as permanently eliminating possibilities, albeit in the name of a more rational form of freedom. This particular bias on the part of the idealists becomes less surprising, once we consider that the philosophy was conceived in the early 19<sup>th</sup> century, prior to the widespread scientific acceptance of the atomic world-view, which only occurred a century later – not least in biology, in which Wilhelm Johannsen expressly coined ‘gene’ in 1905 to refer to an ‘atom of life’. Atomism’s relevance here, especially in biology, is that in principle – and increasingly in practice – it opens the door to a re-engineering of the fundamental elements of matter to produce durable properties and objects that would otherwise not arise spontaneously in nature – either because they would never arise again or they would have never arisen at all. The use of ancient DNA to resurrect extinct species illustrates the former point, the synthesis of new life-forms from novel expressions of the genetic code illustrates the latter (cf. Church and Regis 2012, Schrödinger 1955). In both cases, the path dependency of evolutionary history is effectively subverted, thereby opening up



the future to multiple realizations. Even the mythical status of the unicorn may be reversed by actualizing the relevant atomic combinations to render it as a ‘genetically modified’ horse. Taken to the limit, in the future the Earth itself may be organized as an artificial environment (‘geoengineered’ in the broadest sense) to enable the maximum diversity of species to flourish in what self-described ‘ecomodernists’ call ‘a good anthropocene’ (Nordhaus and Shellenberger 2007).

Thus, the sort of ‘foundations’ to which science allows us access via the atomic world-view may enable us to live in an increasingly *reversible* world. Today, there are already moves along these lines vis-à-vis nanotechnology, whereby chemical agents (‘nanobots’) are released into polluted waters to make them effectively self-cleaning (Fuller 2011, chap. 3). One may take the point still further – into what it will mean to be ‘human’ in the future. According to a certain vision of ‘Humanity 2.0’, genuine human progress is marked by our capacity to live in different worlds, perhaps even in radically different material forms – what transhumanists call *morphological freedom* (Fuller and Lipinska 2014, chap. 3). In that case, our personal identities would migrate across media, say, from carbon to silicon – and perhaps back again? This vision has also attracted the attention of physicists who propose that the very defiance of entropy – i.e. every change of phase state manages to keep at least same number of options open from the agent’s standpoint – may be a marker of intelligence at the cosmic level (Wissner-Gross and Freer 2013).

At one level, this principle revives the efficiency imperative associated with the chemist Wilhelm Ostwald’s failed early 20<sup>th</sup> century attempt to accord energy a metaphysical import that would have normative consequences (Fuller 2000, chap. 2). But at a deeper level, it represents a step in the direction of ‘functionalizing’ our conception of substance, which Cassirer (1923) regarded as the signature revolutionary move in modern thought. In other words, we come to see concrete things not as instantiations of fixed essences but as moments in the confluence of factors in a mathematically defined space. In the first half of the 20<sup>th</sup> century, it was common to speak of this as the ‘Galilean’ break from the Aristotelian world-view. (From this standpoint, Frege is a more ‘progressive’ foundationalist than Husserl.) In that case, one becomes concerned less with maintaining a particular state than a particular dynamic, which may require a periodic change of state, e.g. a quest for the ‘new normal’, which has been the hallmark of ‘progressive’ politics (Fuller and Lipinska 2014, chap. 1). So whereas to be a fully functioning human in 1950 – *à la* American suburbia – may have required a detached house with a front lawn and a garage for the car, in 2050 greater human functionality may be achieved by taking up much less space and consuming much less energy. Moreover, such ‘ephemeralization’, as the visionary engineer Buckminster Fuller memorably called this tendency in

the history of technology, may open one to a greater diversity of ‘normal’ states of being human, simply because less matter and energy would need to be invested in realizing any of them – especially given the anticipated improvements in virtual reality devices.

In the final section, I discuss the philosophical anthropology that is required to recognize humans as beings as capable of providing philosophical foundations. It turns out that the two main approaches to the study of human beings – hermeneutics and naturalism – offer faltering glimpses of the relevant philosophical anthropology, which I will tease out and then develop further, focussing on technology’s role in radically reconfiguring the human condition.

### 3 Recognizing humans as beings capable of philosophical foundations

One of the conceits of the human sciences is that the subjects under investigation normally live their lives as tacit investigators of their own life-worlds. In other words, people are presumed to display a sufficiently high level of epistemic engagement to be considered reliable witnesses, what the sociologist Anthony Giddens (1984) dubbed ‘knowledgeable agents’. Put another way: The highest compliment that academics can pay non-academics is to say that they always already think like academics. Thus, scholarly appeals to empathy and sympathetic understanding (*Verstehen*) have carried a faintly patronizing if not narcissistic quality – especially in an era when people no longer believe that they are descended from the same God in whose ‘image and likeness’ they were created. In the days of Schleiermacher, and possibly even Dilthey and Gadamer, it would have been tantamount to a divine offense for me not to accord you the same degree of perspicacity as I do myself (Schnädelbach 1984, chap. 4). But without a theory of common spiritual descent, the chain of epistemic command goes horizontal: You, dear subject, must be a close observer of your surroundings because I, your investigator, am a close observer my own surroundings – and hopefully yours as well. Various notions like ‘charity’ (Quine) or ‘humanity’ (Davidson), evoking – or parodying? – their religious origins, have been invoked. In any case, the ‘interpretive principles’ cover up the implied methodological egocentrism, whereby you, dear subject, are deemed rational by virtue of basking in my own reflected epistemic glow (Fuller 1988, chap. 6).

What I have just described is, broadly speaking, the ‘hermeneutical’ approach to the human condition. As I have already suggested, it imputes to people a much greater sense of mindfulness than they normally have, as if we behaved

like actors fully conscious of the scripts in which we variously perform. Some of my earliest work was concerned with debunking this imputation – even when applied to scientists – from a broadly ‘naturalistic’ standpoint (Fuller 1993). But more to the point, a relatively unremarked feature of the hermeneutical approach is its implicit delegitimization of the foundationalist philosophical project, including its scientific extensions, *à la* Locke or Russell. (Rorty deserves credit for having emphasized this point.) Hermeneutics presupposes that knowledge claims are ultimately grounded in the claimant’s life circumstances. While this may seem like a truism, the word ‘ultimately’ discounts the possibility that claimants might have access to some deeper reality, be it characterised in causal or transcendental terms. Put bluntly, the very idea that people may know, in some respects, more about times and places they have not directly experienced than the ones they have experienced is a hermeneutical non-starter. Thus, people are rendered ‘transcendental dopes’ (Fuller 2015, chap. 6, esp. p. 241).

Here hermeneutics betrays its own roots in historical-critical theology, with its radically demystified view of religious language as ‘symbolic’ in its referential capacity. Whatever else ‘symbolic’ might mean in this context, it certainly means ‘non-literal’. In the shadow of late imperial anthropology, this non-literalism was read positively as implying that the natives’ beliefs were *not false*, but with the ascendancy of science and technology studies it was read negatively as implying that the scientists’ beliefs were *not true*; hence the ‘Science Wars’ of the 1990s (Fuller 2006). In any case, the epistemic neutrality (or ‘symmetry’) of hermeneutics reflects the field’s own, often failed, diplomatic efforts to chart a course for theological scholarship that does not interfere with the pastoral mission of the churches (Collins 1998, chap. 12). Marx’s *The German Ideology* may be understood as an attempt to derive some profound lessons from one such failure, resulting in the original ‘science war’, fought in Marx’s early adulthood, which focused on the political responsibility of academic theology. One conclusion that Marx (rightly) drew was that if we can indeed talk and act sensibly vis-à-vis a world that lies beyond our normal existential horizons, hermeneutics is incapable of discovering it. Marx was a foundationalist.

In the history of the human sciences, experimental psychology is normally presented as an anti-hermeneutical ‘naturalistic’ approach to the human condition. However, it is easy to forget that experimental psychology began as the psychology of the scientist, understood as the ultimately self-conscious human subject (Fuller 2015, chap. 3). Indeed, until the advent of behaviourism in the early 20<sup>th</sup> century, virtually all of the subjects used in psychology experiments outside of strictly medical settings were scientifically trained people, typically the psychologists themselves. The classic history of experimental psychology (Boring 1950) highlights astronomers’ scrupulous recording of the ‘personal equation’

in observational error as an inspiration for Gustav Fechner to conceive of the study of the mind as the human response to apparatus, aka ‘perception’. Indeed, much of the early history of statistical inference turned on whether variation in response to a fixed stimulus – be it the product of a natural or an artificial environment – should be interpreted as itself normal or, rather, deviations from an implied ‘natural norm’ (Porter 1986). Moreover, whereas astronomers’ training of their telescopic vision reasonably promotes more accurate observations, thereby supporting the idea of a ‘natural norm’, it is far from clear that the same applies to psychological subjects trained to respond more effectively to specific apparatus-based protocols (Hatfield 1991). Indeed, behaviourists – and such early philosophical admirers as Bertrand Russell (1927) – discredited the original psychological method of ‘introspection’ (as this controlled inspection of one’s own mental processes was called) for producing experimental ‘artefacts’: i.e. results that were valid *only* in the laboratory setting.

In their rather different ways, both the hermeneutical and naturalistic approaches to the study of humans turn the subjects under study into an extension of the investigator’s own epistemic horizon. In the hermeneutical case, principles of interpretive charity and humanity can be seen as an abstract version of the sorts of arguments that were used to justify the ‘civilizing mission’ of imperialism in the late 19<sup>th</sup> and early 20<sup>th</sup> centuries. After all, had it not been assumed that apparent world-view differences are surmountable through adequate communication, education, etc., the disruption and violence that often resulted from European expansion into Asia and Africa would have been taken much more at face value – and hence would have probably ended much earlier. However, as long as the imperialists believed that the resistance they encountered simply reflected a misunderstanding between two groups of people who ‘always already’ share the same fundamental beliefs and perhaps even desires, they felt justified in their persistence – especially since the natives were by no means uniformly resistant and often improved their living conditions by adopting the Europeans’ self-avowed ‘objective’ standards.

In the case of the naturalistic approach, the situation is a more complex but also more interesting from the standpoint of what might be called a ‘Humanity 2.0’ style of foundationalism (Fuller 2011). However, a certain conception of ‘naturalism’ needs to be set aside at the outset – one which started to become prominent in the late 19<sup>th</sup> century in the wake of Darwin’s demonstration of the continuity of human and animal life. It is that science is simply an extension of common sense, perhaps a more self-critical application of ‘animal faith’, as George Santayana fashioned it. This position, usually associated with Pragmatism (especially John Dewey’s account of the scientific method but also carried over into Quine), underestimates – in a way Francis Bacon never did – the profoundly transformative role

of technology in the constitution of who we are and our relationship to nature. In most general terms, the increased instrumental mediation of our most valued way of knowing changes the world in which knowledge is made possible – not least because we will have changed in the process; hence, ‘Humanity 2.0’.

In this spirit, the claim of ‘artefactuality’ that led psychologists to abandon the introspective method shortly after the First World War acquired a second life as a thesis in philosophical anthropology: namely, that technology is the extension of the human senses, in which case it may be desirable for us to become more attuned to our apparatus, especially if it is capable of making us more sensitive to the larger world than we would otherwise be, just left to the sensory organs of our birth. In short, a liability of the laboratory – artificially induced experience – came to be recast more positively once apparatus was seen as enhancing, rather than distorting, human capabilities. Nowadays, this argument would be read as a transhumanist defence of our becoming cyborgs. But it can already be found in that great American populariser of German idealism, Ralph Waldo Emerson. Emerson (1870) saw the US Patent Office as building a national exo-physiology – what by the mid-20<sup>th</sup> century would be called a ‘superorganism’ or a ‘noösphere’ – which organizes the products of human inventiveness to maximize their interconnectivity and fertility, the inexhaustible source of prosperity promised to beings whose creativity is underwritten by having been created ‘in the image and likeness of God’ (Fuller 2012, chap. 3). Even before Marshall McLuhan (1965) adopted Emerson as his patron saint – with the television replacing the telegraph in his techno-cosmology – Emerson’s original vision was systematically developed by the German founder of the philosophy of technology, Ernst Kapp, who lived in the United States when Emerson’s influence was at its peak in the mid-19<sup>th</sup> century (Brey 2000).

Not only McLuhan’s media theory but also cybernetics and contemporary transhumanism represent developments in this line of thought. To be sure, the philosophical anthropologists who first charted this intellectual trajectory after the First World War were its opponents, notably represented by Oswald Spengler and Arnold Gehlen. They stressed the dehumanization that allegedly results from the calibration of humanity’s life rhythms to society’s increasing mechanization. It was in this context that the concept of ‘sensory overload’ was first introduced (Fuller 2012, chap. 5). But such critiques overlooked that the ‘machine worldview’, so to speak, opens up new avenues of thought and being, even as it perhaps loses some of the older ones. In this respect, the distinction that bioethicists commonly draw between ‘restorative’ and ‘enhancing’ medical treatments is misconceived for humans living in a technologically extended state – as ‘cyborgs’, if you will. Very few prosthetic devices ‘restore’ to the body to some prior ‘natural’ state. Rather, they compensate for what cannot be ‘restored’, in

any strict sense, while at the same time opening the door to the development of new capacities not available to the natural organism. This is the spirit in which Neil Harbisson started the Cyborg Foundation in 2010 to facilitate the voluntary transition of humans to ‘cyborg’ status, a version of ‘Humanity 2.0’ that is likely to appear more attractive in the future – even among those ‘Humanity 1.0’ beings not especially in need of prosthetics (Wittes and Chong 2014).

When the history of the hermeneutical and naturalistic approaches to the study of human beings are taken together, it becomes clear that the initial impulse to access our untapped potential – namely, that we have the capacity to be more than what is already visible to the naked eye – was lost in its institutionalization. On the hermeneutical side, this potential had been clearly represented by the theological notion that humanity’s recovery from the Fall involves our developing an ability to see the divinity in others, which in secular terms amounts to the recognition that we are meaning-making creatures who know more than may be apparent from our words and deeds; hence, the concept of ‘hermeneutical depth’. However, as hermeneutics became more secularised and academicised in the 19<sup>th</sup> and 20<sup>th</sup> centuries, this impulse was sublimated in more inward notions of ‘reflexivity’ which resulted in an ultimate scepticism about meaning (*à la* Derridean deconstruction) rather than any sense of transcendence. On the naturalistic side, the situation has been more hopeful. Here our untapped potential relates to hidden powers of mind that may be elicited under the right conditions. To be sure, it is a strategy that has inspired a variety of experimental interventions from pedagogy (including IQ tests) to psychiatry (including biomedical treatments) which have been deployed just as often to inhibit as to enable human thought and action. However, an honest appraisal of the situation must include the caveat that what counts as ‘inhibition’ and ‘enablement’ is relative to the envisaged range of appropriate end states for a decent human existence.

In any case, in both the hermeneutical and naturalistic approaches, the empirical study of human beings has always hinted at means of self-transcendence. As people have become self-conscious of how they have been restricted – be it by historical context, genetic programming or simply laboratory conditions – they have also learned what lies on the other side of those restrictions, which in turn enables them to survey the full range of possibilities available prior to action in the future. In this context, technology plays a decisive role in two ways, which are captured by two systems-theoretic conditions: *equifinality* and *plurifinality*. The former refers to alternative means to the same ends, the latter refers to the same means resulting in multiple ends. In the history of biology, these conditions were represented, respectively, by the ‘preformationist’ and ‘epigeneticist’ perspectives on development (Fuller 2012, chap. 2). When applied to the history of technology, they constitute two phases by which humans come to a greater sense of their

own morphological freedom, which provides – so to speak – an ontological clearing necessary for a ‘Foundationalist 2.0’ approach to philosophy. In the first phase, technology shows how the same ends might be achieved more efficiently by humans alienating (we would now say ‘offloading’) their functions to artefacts which they then incorporate into a more comprehensive sense of the human life-world. In the second phase, the inherent powers of those artefacts are exploited to open up new possibilities for human development (e.g. as cyborgs). The result is a foundationalism that is fit for the sort of ‘extended mind’ beings that ‘Humanity 2.0’ is already becoming (cf. Clark 1997).

## Bibliography

- Boring, Edwin G. (1950): *A History of Experimental Psychology*. 2<sup>nd</sup> ed. (Orig. 1929). New York: Appleton-Century-Crofts.
- Brey, Philip (2000): “Theories of Technology as Extension of Human Faculties”. In: Carl Mitcham (Ed.): *Metaphysics, Epistemology and Technology*. Vol. 19. London: Elsevier/JAI Press.
- Cassirer, Ernst (1923): *Substance and Function*. (Orig. 1910). La Salle IL: Open Court Press.
- Cassirer, Ernst (1950): *The Problem of Knowledge: Philosophy, Science and History since Hegel*. New Haven CT: Yale University Press.
- Church, George and Regis, Ed (2012): *Regenesis*. New York: Basic Books.
- Clark, Andy (1997): *Being There: Putting Brain, Body and World Together Again*. Cambridge MA: MIT Press.
- Collins, Randall (1998): *The Sociology of Philosophies*. Cambridge MA: Harvard University Press.
- Descombes, Vincent (1980): *Modern French Philosophy*. Cambridge UK: Cambridge University Press.
- Elster, Jon (1979): *Ulysses and the Sirens*. Cambridge UK: Cambridge University Press.
- Elster, Jon (1984). *Sour Grapes*. Cambridge UK: Cambridge University Press.
- Emerson, Ralph Waldo (1870): ‘Works and Days’ in *Society and Solitude*. Cambridge MA: Welch, Bigelow and Co.
- Fukuyama, Francis (1992): *The End of History and the Last Man*. New York: Free Press.
- Fuller, Steve (1983): “French Science (With English Subtitles)”. In: *Philosophy and Literature* 7, pp. 1–14.
- Fuller, Steve (1985): *Bounded Rationality in Law and Science*. Ph.D. Department of History and Philosophy of Science: University of Pittsburgh.
- Fuller, Steve (1988): *Social Epistemology*. Bloomington IN: Indiana University Press.
- Fuller, Steve (1993): *Philosophy of Science and Its Discontents*. 2<sup>nd</sup> ed. (Orig. 1989). New York: Guilford Press.
- Fuller, Steve (2000): *Thomas Kuhn: A Philosophical History for Our Times*. Chicago: University of Chicago Press.
- Fuller, Steve (2006): *The Philosophy of Science and Technology Studies*. London: Routledge.
- Fuller, Steve (2007): *New In Frontiers in Science and Technology Studies*. Cambridge UK: Polity.



- Fuller, Steve (2009): *The Sociology of Intellectual Life*. London: Sage.
- Fuller, Steve (2011): *Humanity 2.0: What It Means to Be Human Past Present and Future*. London: Palgrave Macmillan.
- Fuller, Steve (2012): *Preparing for Life in Humanity 2.0*. London: Palgrave Macmillan.
- Fuller, Steve (2013): "Deviant interdisciplinarity as philosophical practice: Prolegomena to deep intellectual history". In: *Synthese* 190 (2013): pp. 1899 – 1916.
- Fuller, Steve (2014): "Generational Change as a Vehicle for Radical Conceptual Change: the Case for Periodic Rejuvenation". In: M. de Beistequi, P. Bianco and M. Gracieuse (eds.): *Care of Life: Transdisciplinary Perspectives in Bioethics and Biopolitics*. (Chap. 14). Lanham MD: Rowman and Littlefield.
- Fuller, Steve (2015): *Knowledge: The Philosophical Quest in History*. London: Routledge
- Fuller, Steve and Lipinska, Veronika (2014): *The Proactionary Imperative*. London: Palgrave Macmillan.
- Giddens, Anthony (1984): *The Constitution of Society*. Cambridge UK: Polity.
- Gross, Neil (2008): *Richard Rorty: The Making of an American Philosopher*. Chicago: University of Chicago Press.
- Hacking, Ian (1975a): *The Emergence of Probability*. Cambridge UK: Cambridge University Press.
- Hacking, Ian (1975b): *Why Does Language Matter to Philosophy?* Cambridge UK: Cambridge University Press.
- Hatfield, Gary (1991). *The Natural and the Normative*. Cambridge MA: MIT Press.
- Kuhn, Thomas (1970): *The Structure of Scientific Revolutions*. 2<sup>nd</sup> ed. (Orig. 1962). Chicago: University of Chicago Press.
- Kusch, Martin (1995): *Psychologism*. London: Routledge.
- Laudan, Larry (1981): *Science and Hypothesis*. Dordrecht NL: Kluwer.
- McLuhan, Marshall (1965). *Understanding Media: The Extensions of Man*. New York: McGraw-Hill.
- Merz, John Theodore (1965): *A History of European Thought in the 19th Century*, 4 vols. (Orig. 1896 – 1914). New York: Dover.
- Nordhaus, Ted and Shellenberger, Michael (2007): *Break Through: From the Death of Environmentalism to the Politics of Possibility*. New York: Houghton Mifflin.
- Popper, Karl (1957): *The Poverty of Historicism*. New York: Harper and Row.
- Porter, Theodore (1986): *The Rise of Statistical Thinking*. Princeton: Princeton University Press.
- Proctor, Robert (1991): *Value-Free Science?* Cambridge MA: Harvard University Press.
- Rorty, Richard (1979): *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
- Russell, Bertrand (1927): *An Outline of Philosophy*. London: George Allen & Unwin.
- Russell, Bertrand (1946): *A History of Western Philosophy*. London: George Allen & Unwin.
- Schnädelbach, Herbert (1984): *Philosophy in Germany, 1831–1933*. Cambridge UK: Cambridge University Press.
- Schrödinger, Ernst (1955): *What is Life? The Physical Aspects of the Living Cell*. (Orig. 1944) Cambridge UK: Cambridge University Press.
- Solomon, Robert (1977). "Husserl's Concept of the Noema". In: F. Elliston and P. McCormick (Eds.): *Husserl: Expositions and Appraisals*. South Bend, IN: University of Notre Dame Press, pp. 168 – 181.



Wissner-Gross, Alexander D. and Freer, C.E. (2013): "Causal entropic forces". In: *Physical Review Letters* 110, 168702 (19 April).

Wittes, Benjamin and Chong, Jane (2014): *Our Cyborg Future: Law and Policy Implications*. Washington DC: Brookings Institution.

Guillaume Fréchette

# Two Phenomenological Accounts of Intuition

**Abstract:** Phenomenological accounts of intuition are often considered as significantly different from, or even incommensurable with most of the conception of intuitions defended in analytical philosophy. In this paper, I reject this view. Starting with what I consider to be a relatively neutral phenomenological account of intuition, I first present the main features of Husserl's and Brentano's accounts of intuition, showing the structural similarities and differences between these two views. After confronting them, I finally come back to what unites the two views in order to outline a map of the problem of intuition in which both traditions, the analytical and the phenomenological, appear as two complementary takes on one and the same problem.

## 1 The philosophical dignity of intuitions

A widely-accepted view in the more or less recent philosophical literature on intuitions is that intuitions are a kind of mental state in which *a proposition seems* – or is *presented as* – *true*. Intuitions in this general sense would be a specific attitude towards a proposition, such as:

- i) it is impossible to believe both p and non-p
- ii) it is right to help people in need
- iii) Kirchberg is in Austria
- iv) etc.

Another usage of the term 'intuition' that is also quite common applies it not to mental states, but to the objects of such states. Then propositions like i) to iii) also qualify as intuitions insofar as they are the object of the corresponding attitude. At this point the question arises how one might characterize this attitude apart from describing it as the attitude involved in an intuition. One might describe it as a state of believing, or as an act of forming a belief without inference, or as a disposition to believe. In addition, the following accounts of the attitude involved in intuition have been offered in the past years:

- i) *seeming* (Bealer 1998)
- ii) *intellectually seeming* (Huemer 2001; 2005)
- iii) *being pushed by* (Koksvik 2011)

- iv) *being presented by* (Chudnoff 2013)
- v) etc.

This analysis of the structure of an intuition matches its usual presentation in the literature as a piece of evidence or source of knowledge that *p*. Thus the philosophical dignity of intuitions is often, if not essentially, measured by its epistemic credentials. This would explain the *pro* and *contra* positions discussed in DePaul and Ramsey (1998), where ‘intuitionists’ like Bealer (1998) argue for a specific sense of philosophical intuition distinct from the empirical sciences, and where ‘anti-intuitionists’ like Gopnik and Schwitzgebel (1998) characterize intuition as the result of not fully explicit consciously-observable reasoning processes, making it an instable and thereby suspect source of knowledge, if it has justificatory force at all.<sup>1</sup>

In this context, if intuitions have any place in philosophical discourse, they must have some epistemic credentials that other attitudes lack, otherwise there would be no point in distinguishing intuitions from these other attitudes. Since having epistemic credentials is a property that is also shared by propositions, theories, beliefs, or sets of beliefs, it is natural to think that intuitions are a kind of propositional attitude.

As I have pointed out, if intuitions are to be epistemically efficient, they must act as evidence for something, either in the form of a *prima facie* support (the intuition that sugar is soluble is confirmed by experiment) or in the form of defeasible evidence against a theory (the intuition that Mary learns something when she comes out of the room in black and white contravenes the theory that Mary knew all physical facts). In this sense, having an intuition that *p* is our (at least *prima facie*) justification for believing that *p*. Here again, if intuitions are justifications for our beliefs, they must fit the propositional form of the latter.

These two very general considerations seem to be presupposed by most advocates of intuition in recent debates, to the extent that for most philosophers dealing with intuitions, the use of the term ‘intuition’ in a non-propositional context is simply a derivate use of the expression, which bears only superficial similarities with the philosophical concept, and which may be discarded from philosophical analysis for that reason. Here are some typical examples expressing this view: “Intuitions always take propositions as their objects” (Pust 2012); the view that one could intuit objects concerns a “different phenomenon” altogether (Koksvik 2011, p. 19); basically, all philosophical views of intuition agree that “intuition has representational content” (ibid.), where “representa-

---

<sup>1</sup> See Cummins (1998).

tion” means an explicit conceptual representation as one typically associates it with propositional attitudes; “no extant philosophical account of the nature of intuition contests this assumption [that intuition has conceptual or representational content]” (Koksvik 2011, pp. 7–8). Since often no distinction is made between representational and propositional content, there would seem to be no “extant philosophical account” of the nature of intuition that doesn’t assume that these are propositional attitudes.

Phenomenologists like Brentano or Husserl, on the other hand, do not generally measure the philosophical dignity of intuitions by their justificatory power and do not see them as mental states typically carrying a propositional content. In fact, most phenomenologists would deny this for at least some intuitions. According to them, it is not necessary for intuitions to have a propositional content, or to have justificatory power. However, I don’t take this to mean that there is a deep gap between the phenomenological and the analytical conceptions of intuition, or that these conceptions are (phenomenologically or analytically) incommensurable, either for ideological or methodological reasons.<sup>2</sup> What I would like to suggest is that there is an advantage in thinking of these perspectives as complementary. This advantage becomes more evident when one considers the diversity of phenomenological accounts of intuition, particularly the accounts of Brentano and Husserl. Thinking of both of these accounts as ‘phenomenological’ in some way leads to a more comprehensive account of intuitions which, in our view, is preferable to incommensurabilist views.

In what follows I shall first present the main lines of what I consider to be a relatively neutral phenomenological account of intuition. By ‘neutral’, I mean that features of this account match various conceptions of phenomenology—such as descriptive psychology, metaphysical realism, and transcendental idealism. Second, I will detail the main features of Husserl’s and Brentano’s accounts of intuition. I will show that there are many structural similarities between these two views. After confronting them, I will finally come back to what unites the two views in order to outline a map of the problem of intuition in which both traditions, the analytical and the phenomenological, appear as two complementary takes on one and the same problem.

---

<sup>2</sup> Incommensurabilist views among phenomenologists are rarely presented in these terms. Such a view is to be found in Sokolowski (1981). We also find a similar view in Cobb-Stevens (1990). For a recent incommensurabilist account of intuition, see Wiltsche (2015).

## 2 The Unity of Phenomenological Accounts of Intuition

Before distinguishing the two different accounts, let me first identify the common grounds, from (a) to (e), on which they are based.

a) *Intuition is the most basic form of experience in phenomenology.* Following Husserl in his introduction to the second volume of the *Logical Investigations*, phenomenology works according to an “intuitive procedure” (*intuitives Verfahren*) (Hua XIX/1, p. 6). Following Brentano, it is the paradigmatic case of presentations, which are the most basic mental states.

b) *Intuition is immediate.* When I see the Eiffel Tower before me, or as Husserl says, in *propria persona* (LI V, §2, §11, §14: Hua XIX/1, pp. 356 ff.; 384 ff.; 394 ff.), something is given to me in a direct, immediate, and unmediated way. Brentano characterizes this immediateness in terms of “proper presentation.” According to him, intuitions are proper presentations of sensory content. A proper presentation is a presentation of which the parts of the content are *real attributive* parts.

c) *Intuition is akin to perception.* Phenomenologists like Brentano and Husserl generally characterize intuition (*Anschaung*) as a special form of perception, in opposition to intuitive beliefs, or inclinations to believe, or some other belief-like propositional attitude. In this sense, for Husserl and Brentano, intuitive acts are perceptual, although what counts as perceptual goes over and above sensory perception (for Husserl).

d) *Intuitions are not infallible.* Both Brentano and Husserl agree that the intuitivity of an act doesn’t warrant the infallibility of the perception. If I see what I take to be a woman in a wax museum, my sensory intuition is basically the same, whether I see it *as a wax figure* or *as a real person*. Brentano argues, similarly, that one can have an intuition of a red square when confronted with a different colored geometrical figure.<sup>3</sup>

e) *Intuitions fulfill (Husserl) or satisfy (Brentano) our meaning intentions (Husserl) or our interest (Brentano).* For Husserl, intuition plays a central role in knowledge in virtue of its capacity to fulfill what requires or admits of fulfillment. According to him, perception is an interpretation of intuitive contents. These contents

---

3 In Brentano’s view, intuitions are in normal cases accompanied by an interpretation (*Deutung*) thereof. Cases of optical illusions best illustrate this: in the case of the Müller-Lyer illusion, the intuitions at the basis of my conviction that the lines are unequal are not faulty in themselves – the lines do appear unequal to me not because I would not look at them correctly, but because the content of perception is interpreted incorrectly. On this case, see Brentano (1892; 1893; 1894). On similar cases where no optical illusions are involved, see Brentano (1959, 83 ff.).

fulfill, to a greater or lesser degree, the meaning intention. In Husserl's terms, my act of assertion expressed by "crows are black" has a meaning intention. The assertion involves concepts – the concept of a crow, the concept of something black – that constitute the meaning (*Bedeutung*) of the intention and would be the same in your act of asserting "crows are black." The idea that intuition has a 'fulfilling' function comes from an idea developed by Brentano in his Vienna lectures, which Husserl attended, according to which interest, a mental operation distinct from mere presenting, is satisfied or fulfilled by intuitions and their parts. Thus Husserlian intentions, insofar as they can be fulfilled by given intuitions, are just like the phenomena of 'interest' (*Interesse*) in Brentano, where this interest is always directed at a specific part of the fulfilling content.<sup>4</sup>

### 3 The specific accounts: Husserl

Now that the general phenomenological concept of intuition has been sketched, let me turn to the elements specific to Husserl's account of intuition.

*a. Intuition is not only sensory.* It is widely known that Husserl's concept of intuition doesn't attribute any priority to sense perception or empirical intuition. Intuition comes in many varieties: there are sensory intuitions, like hearing a tone or seeing a color, but there are also intuitions of generality: by using a method of variation, for instance, I can have the intuition of a triangle in general. In this case, the sensory intuition of a color is accompanied by the eidetic intuition of the triangle, a general sense of "seeing," "seeing in the universal sense as an originally presentive consciousness" (*Ideas*, §19: Hua III/1, p. 41 ff.).

*b. Intuition is immediate experience (consciousness).* Husserl uses the term "intuition" or its adjectival form "intuitive" to characterize the immediateness of an experience. How should we understand immediateness? One way of doing so would be to conceive of our experience as it is given to us before

---

<sup>4</sup> On Husserl's early discussion of fulfillment in relation to interest, attention, or pleasure in noticing – a terminology borrowed from Brentano and Stumpf, see Hua XXII, and particularly the *Psychologische Studien* of 1893, where he describes the relation between representation and intuition as a relation between "Meinung und Erfüllung" (Hua XXII, p. 407). When we go from the intuition to the representation, we experience "wie der Mangel beseitigt, wie die Leere ausgefüllt, wie die Hemmung gelöst, die Unbefriedigung zur Befriedigung wird" (Hua XXII, 292). The fulfilling of the intention is a "Lösung der Spannung" (Hua XXII, pp. 407, 411), a "Befreiung," "Erlösung der Hemmung" (Hua XXII, p. 296), Entlastung (Hua XXII, p. 415); but he also describes *Erfüllung* or fulfillment as a "rein im Gegenstand aufgehende Lust am Bemerkten" (Hua XXII, pp. 293, 411). On Brentano's description of interest and its satisfaction, see the Vienna logic lectures (EL72), pp. 12310 ff.

and beyond any interpretation that we usually bestow upon it. Take as an example a particular experience (of mine) of drinking coffee: coffee – you might also call it a medium-dark roast of a specific blend – is “immediately experienced” in the same basic and “immediate” way as my thinking that it will be raining tomorrow. The adjective “immediate” on this reading serves the purpose of underlining the fact that the experience has not been reflected upon: immediate experience is experience “on the go”, as it happens. Immediacy is a feature equally present in all intuitions – it is not *derived* from sensory experience. I experience the coffee “immediately” not because of the felt bitterness when I taste it – although the bitterness is also experienced immediately – but simply because “the coffee” is the object of my experience. All objects of experience – sensory or not – are experienced immediately.<sup>5</sup>

c. *Intuition has a fulfilling function.* In the *Ideas* and elsewhere, Husserl often uses the expression “originary giving intuition” (*originär gebende Intuition*). The “giving” feature mentioned here characterizes what is usually called fulfillment or the fulfilling function of intuitive acts. Let’s come back to the experience of the coffee. My act of thinking the content of the proposition expressed by “the coffee is a Kenyan blend” is what Husserl calls a signitive act. As such, taken in isolation, it is an empty intention. In my experience of tasting the coffee and thinking about it, we could distinguish (among other things) between two moments: my empty signitive act of thinking “the coffee is a Kenyan blend” and the fulfilling intuition, i.e. my experiencing the qualitative properties of Kenyan blends, e.g. the strong acidity, the bitterness, and the tartness of the blend, which make it different from a Brazilian blend, for instance. An intention may or may not be optimally fulfilled – there might be some non-Kenyan beans in the blend, for instance. In this sense, fulfillment is not an “all or nothing” business; it rather comes in degrees. My empty intention, the thought of the proposition “A is red,” may thus be partly fulfilled by an intuition of a green A (see Hua XIX/2, p. 562).

d. *Intuition has evidence.* “Fulfillment” is a functional characterization of what an intuition is. An intuition may or may not fulfill, to different degrees. But these different degrees of fulfillment do not correspond to the different kinds of evidence involved in intuition. Suppose I am thinking of a specific china cup full of coffee. My signitive intention is only fulfilled when I am presented intuitively with this specific cup. But a cup, like all three-dimensional ob-

---

5 The immediateness of intuition does not mean that all intuitions are thereby *originary*. Your intuition is still immediate even if you simply recollect something, or if you imagine something. The distinction between originary and non-originary does not pertain to the immediateness of intuition.

jects, is always given under some perspective. The problem here is that perceptual intuition, even when it gives me all that it can (I perceive the cup under the best lighting conditions, from a very short distance), is still a structurally limited kind of evidence. It is open to revision at a later stage of experience. In an ideal situation I would be given perfectly adequate evidence of the cup, but as in all normal perceptual cases, evidence is only more or less adequate. My thinking that the three angles of a triangle sum up to 180 degrees is fulfilled by the eidetic intuition of triangles. But in this case there is no degree of fulfillment. The evidence characterizing such a type of fulfillment is called apodictic evidence.

To put it simply, evidence is a *property* of intuitions, of which fulfillment is a function. It is thanks to this property that intuitions are said to provide pieces of knowledge. To know that the knife is on the table is to have a meaning intention (expressed by the sentence “the knife is on the table”) *fulfilled* by the corresponding intuition.

## 4 The specific accounts: Brentano

*a) Intuition as the paradigmatic case of presentation.* Brentano’s philosophy of mind takes intuition as its most fundamental element. In his architecture of the mind, presentations constitute the most basic components. These components form the basis of a second mental category: judgments; and for a third as well: acts of love and hate, or, more generally, emotions. Contrary to judgments and acts of emotions, presentations have no valence. However, there are different kinds of presentations, distinct from one another by degree of clarity: the purest form of presentation, as we may call it, is intuition (*Anschaung*). They have a high degree of clarity: the colors I see and the tones I hear are, in Brentano’s sense, intuitions. Brentano’s examples in his lectures on aesthetics, which Husserl also attended, are the following: I have an intuitive presentation when I see a red square. I have an unintuitive presentation when I think of a round square. But I also have an unintuitive presentation when I think of a red square.

In some sense, one could say that all presentations are either intuitions or are based on (in the sense of being partly constituted by) an intuition. Brentano holds a similar thesis regarding mental activities more generally: they are either presentations or based on presentations.

*b) Intuition vs imagination.* What happens when I first see a red square and then imagine a red square? According to Brentano, the first act is a pure intuitive presentation, while the second one is unintuitive, although it has some degree of intuitivity. How is this to be understood? Brentano, relying on Aristotle, says that the content of these acts is different, although they *seem* to have the same



content.<sup>6</sup> Only in an improper (*uneigentlich*) way – i.e. in the way they *seem* to be – is it correct to say that they have the same content. But in fact, the imagination has as its content a thought with an intuitive kernel. Accordingly, intuition and imagination are not distinct *modes* of presentation, but differ in virtue of the properties of their contents: intuitions, for instance, have a degree of intensity that non-intuitive presentations lack.

c) *Intuition and attention*. Like Mill, Brentano rejects the idea that there are general concepts. Acknowledging only one mode of presenting, of which intuition is the paradigmatic case, Brentano proposes an alternative account of how we represent things or properties like triangularity. In his Vienna lectures on logic of 1884/85 (mentioned earlier), Brentano explains his theory in the following way:

[I]n relation to the question of universals, it appears that when I also have no other presentations than individual presentations, in a certain way, I do have them [i.e. universals] – namely as partial presentations circumscribed through a particular interest – and this way is sufficient to give to the general name not simply a plurality of equivocal individual meanings, as the nominalists wanted, but rather a unitary, truly general sense. (Brentano EL 72, p. 12349)

As such, there is a sense in which we can say that Tom, Dick and Harry all form the *same* general concept of red, provided that they focus their attention on the same features of some presented object. In this way, the abstract term “color” is not a simple fiction. Instead, it has a “truly general sense”, without requiring the acceptance of an abstract entity corresponding to the term ‘color’.

This account introduces general presentations as having *abstracta* as intentional entities, isolated on the basis of an act of interest. These acts of interest take as their objects some parts of a presentation contents that are not intuitive as such. In other words, redness as an object of presentation is constituted by intuitive and non-intuitive parts: the intuitive parts are the visual content or individual presentations and its properties (hue, brightness, constancy, etc.), while the non-intuitive parts (the property of being a colour) are co-present in the presentation content, but are not accessible in presentations as such. Thanks to an act of interest in the relevant part of the presentational content, we can isolate the abstract presentation, which otherwise would simply be an indistinct part of the intuitive presentation of the red colour. Partial presentations focus upon the object of a particular interest (*Interesse*), a term that Brentano uses here as

---

<sup>6</sup> See Brentano (1886: Lecture 33) and (1959, p. 85 ff.).

a German translation of Descartes' notion 'admiratio'.<sup>7</sup> This interest, also called "attention" (*Aufmerksamkeit*) is often characterized as a pleasure in noticing (*Lust am Bemerken*). Concepts are then defined as partial presentations that may serve as mediators for further psychical activities, but they are concepts on the sole basis of the act of interest directed towards them. No parts of presentations are intrinsically conceptual; they are made conceptual by an act of interest. As Brentano puts it:

[T]here is only one mode of presenting activity, [...] [but] through the detaching and unifying force of a particular interest, directed exclusively upon one or certain parts of the complete presentation, these parts of presentation can become mediators of nomination and the presentational basis of judgments and emotional activities. (Brentano EL72, p. 12340)<sup>8</sup>

Following this account, when I see a red table, the presentation I have of the red table is a partial presentation, which in itself is not intuitive: what I actually see are only patches of colour. But this partial presentation contributes to the individuation of the object of my presentation. This contribution is not effected on the basis of a new mental activity – it's just that this "partial presenting" becomes the object of an act of attention. This attention has the character of a judgment: I see the red table and I judge that 'this presenting shows a red table' or that "this presenting of a red table exists". The focus thus bestowed upon the partial presentation of the table "elevates", so to speak, the partial presentation to the level of a mediator (*Vermittler*) or a sign. According to this account, concepts are to be considered as modified intuitions: they are modified thanks to the focus bestowed upon them by specific judgments or acts of interest or attention.

d) *Intuition as acceptance*: Another important feature of Brentano's conception of intuition is its "indissoluble" connection with acceptance. Brentano's student Anton Marty describes this connection as follows: "the sensation is an act which contains two mutually inseparable parts, the intuition of the physical phenomenon and assertoric accepting thereof."<sup>9</sup>

Intuitions never come in isolated form: they are always indissolubly tied to the assertoric acceptance of what is presented in the intuition. The idea here is that I cannot but accept what is given intuitively: my intuition of red roses cannot occur without my somehow "accepting" the red roses.

This act of acceptance may still be more primitive than the *judgment* that there is a red rose before me. One reason for drawing this distinction is that

---

<sup>7</sup> See EL 72, p. 12315.

<sup>8</sup> Marty (1894/2010, p. 125) uses the exact same sentence.

<sup>9</sup> As quoted from Kraus (Engl. 1976, p. 235).

while in a presentation I can only accept what is presented to me, in an act of judgment I can either accept *or reject* what is presented. Intuitions do not have a valence, for this reason, even if an acceptance is involved in them: there is no possible alternative – the acceptance involved in intuition cannot possibly turn into rejection. Second, my judgment that a rose is before me is based on the fact that I cannot be mistaken that I have a *presentation* of a rose in making this judgement. It is therefore not a ‘blind’ judgement that lacks this inner perception of a presentation on which it is based. The difference between blind acknowledgement, as a primitive cognitive activity, and judgment *per se* as a higher cognitive activity allows Brentano to argue in cases like the Müller–Lyer illusion that we blindly acknowledge the different lengths of the lines, but recognize them as being equal on a higher level of cognitive activity.<sup>10</sup>

## 5 Confronting the Views

As suggested above, Brentano and Husserl share the “neutral” and more general phenomenological view of intuition according to which intuition (a) is a form of or based on perception; (b) is immediate (“present” in Husserl’s terminology; “proper” in Brentano’s); (c) is a source of knowledge; (d) is the basis of imagination; and (e) comes together and is completed with an “empty” intention (a “meinen” in Husserl’s terminology, or an act of attention or interest; a “pleasure in noticing,” in Brentano’s).

These five core features together form the ‘neutral’ phenomenological account. But what distinguishes the two accounts? From what we have discussed, three main differences emerge. First, Husserl and Brentano disagree on the tie between intuition and acceptance. Husserl rejects the idea that presentations include any doxastic element, be it assent or judgment proper. Second, Husserl accepts two kinds of perception, according to which one can also perceive universals. Brentano and the Brentanists reject this.

The third and most important difference between these two positions is that for Husserl intuition is (a mode of) consciousness while for Brentano intuition is a presentation. Husserl often describes the “fulfilling” experience of intuition as a “peculiar *consciousness of fulfillment*” (LI VI, §8, p. 694). In the *Ideas* in particular – and also afterwards when Husserl describes the activity of intuition – he speaks of perceptual consciousness. This is to be contrasted with what Husserl from early on calls “representation” (*Repräsentation*), but also from acts of

---

<sup>10</sup> See for instance Brentano (1987), p. 26.

meaning. Perceptual (fulfilling) acts and meaning acts constitute two fundamentally different categories of acts.

At this point Husserl clearly departs from Brentano. According to the latter, the difference between an intuition and a non-intuitive presentation originates in the nature of the object presented, and not in the mode of consciousness through which the object is given. According to Brentano, intuitions are *proper* presentations, while a surrogate presentation (my presentation of a figure with 1000 sides, for instance) is an improper presentation.

This difference is visible in Husserl's arabesque example. Imagine that you look at some arabesque. First you see it merely as a decorative symbol. At some point, you realize that it represents a linguistic symbol.<sup>11</sup> Or, alternatively, take Fig. 1.



Figure 1

What happens between the moment you see the arabesque (or the smiling face in Fig. 1) as mere decoration or picture and the moment when you realize that it is a linguistic symbol (in this case the word “old”)? “The whole mode of consciousness changes,” says Husserl: “you *see* the sign, but you are not focusing on it, you are not intuiting it” (Hua XXII, pp. 115–116). In some sense, one could say that the seen object is “dimmed” when the switch happens: it remains the same object, the same content of perception, but one’s attitude towards it is different.

Brentano would answer that the only difference is that, driven by some changing interest, you focus your attention on a different *part* of the object. This is the only change that occurs. The object is not ‘dimmed’ as a whole; parts of it are simply ‘shut down’ in order to better focus on one part of it. The explanation is quite simple: the sign that you perceive is ‘in’ the picture, though at all times you continue to see the picture.

If we leave some of these descriptive differences aside, it might seem as if this difference concerning intuition is not particularly important. After all, Brentano speaks of a change of interest, or a focus of attention, where Husserl speaks of a different mode of consciousness. But this change in the mode of consciousness is for Husserl also an indicator of intentionality. For Husserl, when you look at the face, and then look at the word, the content doesn’t change: both have the same *intuitive* content. What changes is the object perceived, and this change is registered in the different ways in which the intuitive content is intended.

<sup>11</sup> See also Hua XIX/1, p. 399; XIX/2, p. 566–7.

Brentano would say that these “ways in which a content is intended” are simply parts of a single whole, but that the way this single whole is perceived doesn’t change. The mode of “intentional relation” remains the same in all cases.

## 6 Remapping the Problem

As we have seen, we arrive at an account of intuition with a far wider scope when we accept the core common theses that unite these two phenomenological accounts of intuition – particularly the first thesis, according to which intuitions are perceptions (and perceptions are not propositional attitudes). Such an approach also allows us to appreciate the complementarity of the phenomenological and the analytic accounts of intuition. Indeed, it allows us to see analytical and phenomenological accounts of intuition branching out from one single root: where analytic philosophers argue for or against the thesis that intuitions are *sui generis* mental states, phenomenologists like Brentano and Husserl argue for or against the thesis that intuition is a particular mode of consciousness. This branching out could be illustrated as in Fig. 2

According to the analytic account of intuition discussed above, only the two right branches of the schema have philosophical dignity, while for at least some phenomenologists, the two left branches are philosophically relevant. My discussion of Brentano and Husserl in sections three to five showed that basically the same distinctions are made in phenomenology and analytic philosophy when intuitions are classified either as presentations or as beliefs or as a *sui generis* mode of consciousness or a *sui generis* mental state. In both cases, the distinction covers a range of answers to the question of whether intuitions form, in themselves, a category of mental states. On this point, advocates of the view that intuition involves an intellectual seeming and defenders of a categorial intuition in phenomenology both answer in the affirmative. Whether this kind of mental state is essentially propositional and/or perceptual is at this point secondary. The important point is that the complementarity view advocated here allows us to use the resources of one sibling node (e.g. phenomenological description) to address issues from the other sibling node (e.g. conceptual analysis), and vice versa. Analytic philosophers have started to acknowledge this kind of complementarity.<sup>12</sup> It is to be hoped that phenomenologists will do likewise.<sup>13</sup>

---

<sup>12</sup> For a recent contribution in this direction, see Chudnoff 2013.

<sup>13</sup> This paper was written with the support of the Austrian Science Fund (FWF: Project M1403-G15). Thanks to Johannes Brandl for helpful comments and suggestions.

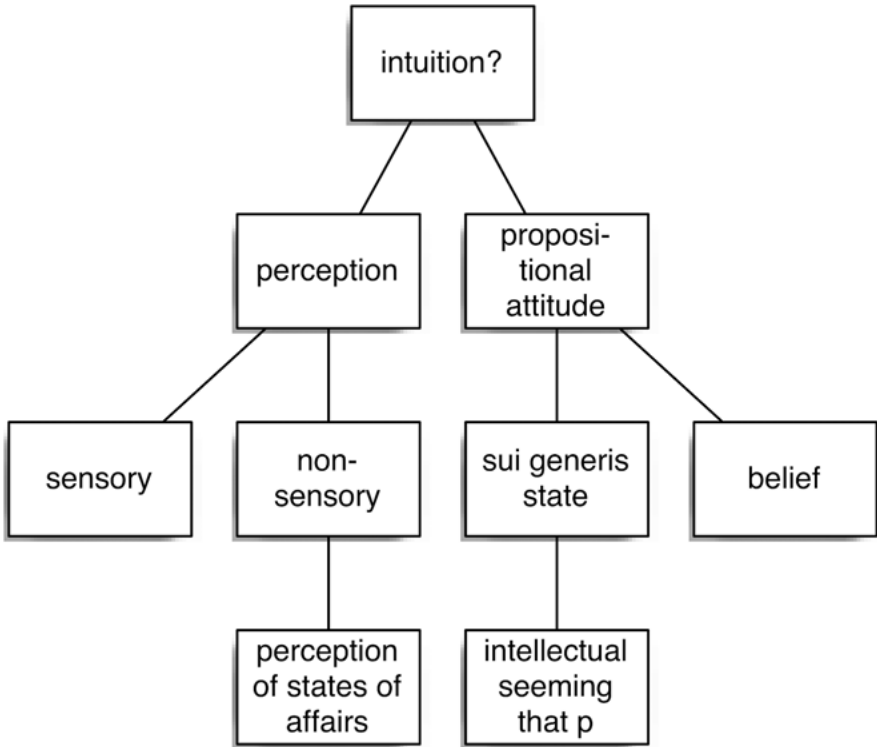


Figure 2

## Bibliography

- Bealer, George (1998): "Intuition and the Autonomy of Philosophy". In: DePaul and Ramsey 1998.
- Brentano, Franz (1892): "Über ein optisches Paradoxon". In: *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* III, pp. 349–358.
- Brentano, Franz (1893): "Über ein optisches Paradoxon. Zweiter Artikel". In: *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* V, pp. 61–82.
- Brentano, Franz (1894): "Zur Lehre von den optischen Täuschungen". In: *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* VI, pp. 37–45.
- Brentano, Franz (1884/85): *Logikvorlesung*. EL72. Manuscript. Houghton Library, Harvard University.
- Brentano, Franz (1886): *Ausgewählte Fragen aus Psychologie und Ästhetik*. Manuscript. Houghton Library, Harvard University.
- Brentano, Franz (1959): *Grundzüge der Ästhetik*. Bern: Francke Verlag.

- Brentano, Franz (1987): "Von der Natur der Vorstellung" (J. Brandl, ed.). In: *Conceptus* 21, pp. 25–31.
- Chudnoff, Elijah (2013): *Intuition*, Oxford: Oxford University Press.
- Cobb-Stevens, Robert (1990): *Husserl and Analytical Philosophy*. Dordrecht, Kluwer Academic Publishers.
- Cummins, Robert (1998): "Reflections on Reflective Equilibrium". In: DePaul and Ramsey 1998.
- DePaul, Michael/Ramsey, William (eds.) (1998): *Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Inquiry*. Lanham, MD: Rowman and Littlefield.
- Gopnik, Alison / Schwitzgebel, Eric (1998): "Whose Concepts are they, Anyway? The Role of Philosophical Intuition in Empirical Psychology". In: DePaul and Ramsey 1998.
- Huemer, Michael (2001): *Skepticism and the Veil of Perception*. Lanham: Rowan & Littlefield.
- Huemer, Michael (2005): *Ethical Intuitionism*. New York: Palgrave Macmillan.
- Husserl, Edmund (1976): *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie* (K. Schuhmann, ed.). Den Haag: Martinus Nijhoff (Hua III/1).
- Husserl, Edmund (1979): *Aufsätze und Rezensionen (1890–1910)* (B. Rang, ed.). Den Haag: Martinus Nijhoff (Hua XXII).
- Husserl, Edmund (1984): *Logische Untersuchungen. Zweiter Band: Untersuchungen zur Phänomenologie und Theorie der Erkenntnis* (U. Panzer, ed.). Den Haag: Martinus Nijhoff (Hua XIX/1 and XIX/2).
- Kraus, Oskar (1976) "Toward a Phenomenology of Time Consciousness". In: Linda McAlister (ed.): *The Philosophy of Franz Brentano*. London: Duckworth, pp. 224–239.
- Marty, Anton (1894/2010): *Deskriptive Psychologie*. (M. Antonelli and J. Marek, eds.). Würzburg: Königshausen und Neumann.
- Pust, Joel (2012): "Intuition". In: Edward Zalta (ed.): *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/intuition/> (last accessed Aug 6, 2014).
- Sokolowski, Robert (1981): "Husserl's concept of categorial intuition". In: *Philosophical Topics* 12, pp.127–141.
- Wiltsche, Harald (2015): "Intuitions, Seemings, and Phenomenology". In: *teorema* XXXIV/3, pp. 57–77.

Julia Jansen

# Imagination and 4E Cognition: An Analytic-Continental Exchange

**Abstract:** The interdisciplinary field of phenomenology and cognitive science is thriving. Recently, it has gained even further ground through the rise of a new cluster of paradigms in philosophy of mind and cognitive science, which can be collectively referred as the model of “4E cognition”. This model takes cognitive processes to be *Embodied*, *Enactive*, *Embedded*, and/or *Extended*. Those 4 e’s enable fertile crossovers between analytic and continental approaches in philosophy of mind, phenomenology, and cognitive science. All of this is very promising indeed. However, despite the irrefutable significance of these new developments, more work needs to be done to ease up constraints and limitations that attest to the lasting effects of “common sense” naturalist, or positivist biases in philosophy of mind and the cognitive sciences. These biases may have little effect on research focused on perceptual cognition, but they pose noticeable constraints on research focused on imagination. In this paper, I identify and then challenge some of these biases with a view to making the model of 4E cognition more obviously applicable to issues of imagination. However, I believe it is important to acknowledge that other obstacles also hamper imagination research in significant ways and for more than merely philosophical or scientific reasons.

The interdisciplinary field of phenomenology and cognitive science is thriving. Recently, it has gained even further ground through the rise of a new cluster of paradigms in philosophy of mind and cognitive science, which can be collectively referred as the model of “4E cognition”. This model takes cognitive processes to be *Embodied*, *Enactive*, *Embedded*, and/or *Extended*. Those 4 e’s enable fertile crossovers between analytic and continental approaches in philosophy of mind, phenomenology, and cognitive science. All of this is very promising indeed. However, despite the irrefutable significance of these new developments, more work needs to be done to ease up constraints and limitations that attest to the lasting effects of “common sense” naturalist, or positivist biases<sup>1</sup> in philosophy of mind and the cognitive sciences. These biases may have little effect on research focused on perceptual cognition, but they pose noticeable constraints

---

<sup>1</sup> I use the term “positivism” in order to point to a cluster of approaches, which also include naturalism, scientism, realism, etc. and which are widespread far beyond academic circles in society at large. I will give specific examples in the second section of this paper.



on research focused on imagination. In this paper, I identify and then challenge some of these biases with a view to making the model of 4E cognition more obviously applicable to issues of imagination. However, I believe it is important to acknowledge that other obstacles also hamper imagination research in significant ways and for more than merely philosophical or scientific reasons. I therefore begin with some observations concerning such general impediments to imagination research and their cultural history with specific reference to the reception of Kant's and Husserl's account of imagination. Second, I turn to contemporary research in 4E cognition in order to show some limitations of that model, which, however, I believe are corrigible. Third, I offer some reflections on a possible inclusion of issues of imagination in the model of 4E cognition, which could, in my view, constitute a genuine exchange between analytic and continental thinking on cognition and consciousness.

## **1 Historical difficulties and some reasons for optimism: Responses to Kant's and to Husserl's accounts of imagination**

At least since the late 19<sup>th</sup> century, Kant scholarship was dominated by epistemological and logical readings that have been, at best, unfavourable and, at worst, outright hostile toward Kant's notion of imagination. Heidegger's powerful interpretation of Kant (Heidegger 2010) is a famous exception but remains just that: an exception that only further proves the rule. When Kant began to attract new interest amongst analytic philosophers in the 1960s (P.F. Strawson 1966; Bennett 1966), older neo-Kantian influences were sustained or replaced by new efforts to salvage Kant's "transcendental arguments" from his flawed metaphysics and his misguided faculty psychology. Both neo-Kantian and analytic readings therefore eclipsed the imagination from their "proper," that is, "properly philosophical" reconstructions in an attempt to purge Kant's philosophy from its "unphilosophical" psychologicistic elements.

Husserl, at least initially, was also very dismissive of Kant's philosophy and especially of his notion of a transcendental synthesis of the imagination. There is nothing surprising about this. Husserl came to philosophy in a thoroughly neo-Kantian climate and formulated his phenomenology, in part, in opposition to it. In the neo-Kantian light of the day, it was almost inevitable that Husserl would regard Kant's philosophy as a prime example of the kind of abstract and formalistic metaphysics that he, Husserl, was working against. The fact that his teacher Franz Brentano considered the Kantian position "decadent" (Körner 1987, p. 11)

would have only added to Husserl's antipathies. Moreover, Husserl's own concerns with psychologism led him to reject Kant's notion of the imagination in particular as "mythical" and "anthropological" (Husserl 1956, p. 228). Husserl's early negative assessment of Kantian philosophy stuck with some phenomenologists after him,<sup>2</sup> even though Husserl himself later revised it substantially – based upon his own comparatively intensive study of Kant's first *Critique* and upon his equally intensive relation to his neo-Kantian colleague Paul Natorp. When Husserl was rediscovered by analytic philosophers in the second half of the twentieth century (Føllesdal 1969; Smith & McIntyre 1975, 1982; Dummett 1993) he was read alongside Frege as a philosopher who was concerned with distinctions between sense and reference and with the truth conditions of propositions; or alongside Brentano as a proponent of a representationalist theory of mind. Other elements of Husserl's work were downplayed, among them most prominently his appropriation of the notion of synthesis, which in his later works he had praised as Kant's "great discovery" (Husserl 1970, p. 103f.). Similarly, Husserl's own account of imagination, which he developed under the title phantasy" (*Phantasie*) as a phenomenologically clarified alternative to Kant's "mythical" notion of imagination" (*Einbildungskraft*) (Husserl 1980), has received little attention to this day (although interest is increasing). Husserl's own claims of the systematic and methodological importance of the imagination and his declaration of fiction as the "vital element" (*das Lebenselement*) of phenomenology (Husserl 1950, p. 132) appear to have caused considerable embarrassment amongst Husserlians and is often left aside.<sup>3</sup> Too heavy, it appears, weigh the numerous criticisms of Husserl's method of "eidetic variation" – which was at least one reason for the vital importance for phenomenology Husserl ascribed to the imagination – as neo-Platonic "metaphysical frippery" (Kraft 1957) or dangerous "irrationalism" (Adorno 1982).

Behind much of the discomfort with the topic of imagination stands an even greater scepticism towards Kant's and Husserl's declared idealisms. In fact, it is hard to find aspects of Kant's philosophy that have caused as much irritation

---

2 For example, much of Maurice Merleau-Ponty's *Phenomenology of Perception* is written against what he took the Kantian (and Cartesian) thesis, namely the thesis of the constitutive power a transcendental apperception (or *ego cogito*): that "I could not apprehend anything as existing, unless I, first of all, experience myself as existing in the act of apprehending it" (Merleau-Ponty 1962, p. ix).

3 Husserl anticipated that his statement regarding the "vital element" of phenomenology would be mocked. In a footnote directly referring to that phrase he remarks that it "would be particularly well suited as a quotation to ridicule, from a naturalistic point of view, the eidetic manner of knowing" (Husserl 1950, p. 132 fn.).

and found as little appreciation as his ideas on idealism. Although this has not always been the case – the idealists and romanticists of the 19th century certainly were enthusiastic aficionados – it certainly has been the case since the early 20th century. The standard rejection of Kant’s idealism thus dovetails nicely with a widespread consensus amongst 20th century and contemporary philosophers about the untenability of idealism in general. Remember also that G.E. Moore’s and Bertrand Russell’s rejections of idealism is considered one of the founding moments of analytic philosophy (Moore 1903). Around the same time, Neo-Kantian philosophers, especially of the so-called “Marburg school” attempted to go back behind neo-Hegelian philosophical excesses in order to retrieve realist, meta-scientific aspects of Kant’s work for a new “reasonable” philosophy. And it is fair to say that Husserl’s phenomenological approach had initially found such eager followers in part because it promised a new realism that would finally put an end to 19<sup>th</sup> century flights of fancy that were seen to have misled philosophy into the quagmire of unjustifiable speculation and metaphysical construction. All the greater was the disappointment amongst Husserl’s disciples, of the Göttingen and Munich schools and beyond, who saw no other way but to turn their backs on him when he himself “came out” as an idealist with his 1907 lectures on *The Idea of Phenomenology*, in which he also introduced the idea of a phenomenological “reduction” (Husserl 1973).

These many coincidences are unlikely to be mere coincidences. The vehement dismissal of Kant’s and Husserl’s idealisms and of their notions of imagination can only become routine in a milieu that – influenced by a general materialistic-realistic *Zeitgeist* – may now have been, under different guises, biased against idealism for over a hundred years.

That said, the field of Kant studies has undergone significant changes. Amongst them, a widely noticed resurgence of interest in Kant’s transcendental idealism as a genuine alternative to pre-Kantian idealisms and contemporary representationalisms (Allais 2015; Allison 2004; Hanna 2006; Langton 1998; Schulting & Verburgt 2011); and an increasing recognition of Cartesian (and Lockean) biases in 20th century analytic interpretations of Kant (Abela 2002). It is therefore fair to say that a “new Kant” is emerging from the most current scholarship. This new, or as Bird has put it, “revolutionary Kant” (Bird 2006) has been accompanied by some steps towards a rehabilitation of Kant’s notion of a “transcendental synthesis of imagination” (Makkreel 1990; Gibbons 1994; Longuenesse 1998; Banham 2006). While this recent development has not yet been officially recognized as a paradigm shift in Kant studies, but is instead being observed by proponents of the “standard picture” as a very interesting and especially lively sub-debate, it might indeed lead to substantial corrections of that picture with regard to questions of idealism as well as imagination in Kant.

In a parallel move, the field of Husserl studies has changed considerably since the late twentieth century. A “new Husserl” (Welton 1998, 2003) has emerged from a line of interpretation that emphasizes Husserl’s later, only relatively recently published and translated, “genetic” works. In opposition to the “old” accepted view that Husserl advanced a representationalist, idealist theory of consciousness, this new reading takes Husserl as concerned not with mental items and their relations amongst each other (as Husserl’s earlier “static analyses” were first taken to suggest), but instead with the processes in which meaning is first of all generated and in which consciousness constitutes itself. For this “genetic” account, the idea of synthesis is indispensable – and so overly rash dismissals of post-Kantian elements in Husserl’s work had to be reviewed. Moreover, this new reading of Husserl has also led to a more positive reappraisal of his project of an expressly transcendental notion of phenomenology (Moran 2008; Zahavi 2002) and even, considerably more reluctantly, to a more sympathetic assessment of his transcendental idealism (Luft 2011; Naberhaus 2007; Bernet 2004). Since the 2005 publication of John Brough’s English translation of Husserl’s lectures on *Phantasy, Image Consciousness, and Memory* (Husserl 2005), there has even been some renewed interest in Husserl’s reflections on the imagination. However, so far, this interest has mostly focused on Husserl’s descriptions of imagination as a specific act of consciousness (i. e., “phantasy consciousness”), and less on the systematic role that the imagination plays in Husserl’s phenomenology.

Old biases are still operative, and perhaps also conscious or unconscious desires to present phenomenology in the most “respectable” realist light, which might be easier if we avoid the idea of a fundamental systematic methodological function of the imagination at the heart of the project. And yet, there are reasons for optimism. The recent developments in both Kantian and Husserlian circles mean that old biases are slackening their grip. This allows for new interpretations of the role of imagination and of related issues concerning transcendental idealism in Kant and Husserl. What’s more, given the interchanges and feedback loops between more continental, historical and more analytic, thematic approaches, the loosening up of those circles also has effects on the contemporary interdisciplinary field of phenomenology, philosophy of mind and cognitive science (and vice versa).

## 2 Current difficulties and some reasons for optimism: “4E” biases

The model of 4E cognition facilitates a lively exchange between continental and analytic philosophical ideas. In fact, phenomenological approaches can be considered (not exclusively, but perhaps most obviously) “antecedents” of 4E cognition, and of situated models more generally (Gallagher 2009). After all, the claims that cognition is in part

constituted by features of embodiment; actively performed and not simply instantiated by a mental or neurological state; embedded in a surrounding environment; and, finally, as an activity, extended not only beyond the brain but often beyond the limits of the body

are much closer to, albeit not identical with, views that are central to continental, and in particular to phenomenological models of consciousness than earlier disembodied, internalist, and computationalist paradigms in philosophy of mind and cognitive science. However, at least initially, this “new science of the mind”, which is said to be “sweeping the planet” (Rowlands 2010), was leaving issues of the imagination behind. Research on the imagination in this newly emerging field has been hampered by biases that are partially grounded in practical limitations of experimental science but also, in my view, obviously facilitated by a general social milieu that endorses and enforces positivistic, scientific, naturalistic views. Amongst these biases are:

1. A lingering indebtedness to *behaviourism*: Note that 4E cognition is precisely a model of *cognition*, not of *consciousness*. By cognition, researchers in cognitive science and in related debates in philosophy of mind, usually mean information processing.
2. Within the dominant *functionalist* framework, cognition is not tied to a specific (human or non-human) processing architecture, but crosses over boundaries between human, animal and A.I. cognition. Researchers are mostly concerned with subpersonal processes associated with adaptive flexible, not necessarily human, behavior.
3. These processes must, of course, be 3<sup>rd</sup>-*person observable*.
4. Moreover, in line with the persisting power of the functionalism model, *intentionality is often regarded in isolation from phenomenal consciousness*. Intentionality is then understood simply as the “aboutness” of, for example, mental or neural states or processes and computational symbols (also referred to as “representations”). Phenomenal consciousness is then attributed to

experience only in so far there is “something it’s like” subjectively to undergo it (i. e., to “qualia”).

5. In this context, philosophical debates have shifted significantly from traditional concerns with representational contents to (scientifically testable) questions about (experimentally observable) *vehicles* of content (for example, neural or algorithmic vehicles).
6. All of this leads, not necessarily intentionally, but effectively to what we might call the “*perception bias*” in 4E cognition (Jansen 2013). For much research is committed to the investigation of “densely coupled” processes that facilitate observable “adaptive flexible” behavior in the “presence of a tangible target”. What is taken to be at stake is “real-world action and intervention”, which involves problem-solving observable according to “task relevant inputs and outputs” (Clark 2005, p. 233f.; Wilson 2002, p. 626).

None of this bodes well for imagination.

Sure, imagination may involve problem-solving (think of thought-experiments, for example) and may be in this sense target-related. But even when it is, and there are a lot of cases of imagining where it isn’t, it need not involve any observable behavior; or at least it is difficult to identify reliably the behavior that correlates with the “mental” exercise of imagining. For the same reason, it is difficult to determine the mental vehicles that “transport” specific episodes of imagining. Further, the imagination *prima facie* seems to resist at least a strongly functionalist approach since even those who have no difficulty in thinking of cognition as a process that may be instantiated in human and non-human minds, are often much more reluctant to “functionalize” the imagination to the same extent. Moreover, imagining is generally seen as involving conscious states, which makes it hard to integrate it in a model that largely deals with sub-personal intentional processes, but that does not necessarily extend to phenomenal consciousness. If we were to think of it as “extended” in some way, this would bring on the idea of phenomenal externalism, that is, the idea that the phenomenal quality of a given experience depends upon more than just states or processes internal to the brain of the experiencer. This may be seen as a far more radical idea than even the idea that cognition is extended in this way, and thus faces even more resistance.

And so it seems to be the tacit conclusion drawn, imagining must be an “internal” mental activity after all, involving good, old-fashioned “internal” representations (Clark and Torribio 1995; Shapiro 2011). It appears, then, that imagination does not qualify for integration into the new field of research. There is, so it seems, no 4E imagination.

However, this is not the last word on the matter. Phenomenologists have rich resources at their disposal to correct common ideas about the imagination – resources offered, for example, by Husserl (2005) and Sartre (2004). Moreover, recent research in philosophy of mind also reveals a richer and much more refined account of imagination than the one tacitly assumed in early 4E cognition. Far from being an “off-line” internal viewing of mental imagery, largely disconnected from the “actual goings-on” of practical life, imagination is increasingly seen as central to important modes of knowing and as critical for access to the modal articulation of cognition, which cannot be reduced to actual events in the immediate, perceptual environment of a cognizing system (Kind 2016; Kind and Kung 2016). These new insights into the complex nature of imagination, if introduced into the model of 4E cognition, are bound to lift some of the constraints hitherto imposed on applications of that model to cases of imagination.

### 3 4E Imagination? An Analytic-Continental exchange

It is not possible here to deliver a new comprehensive account of imagination that integrates phenomenological resources and recent analytic research, and also to demonstrate how this account leads to an inclusion of imagination under the model of 4E cognition. I therefore restrict myself to a few pointers, which, however, in my view quickly show that the idea of a 4E imagination cannot easily be dismissed.

First, even just a look at the phenomenological account of the imagination offered by Husserl, already challenges the common internalist model, which leaves imagination “representation-hungry” (Clark and Torribio 1995). Husserl proposes a notion of imagination (“phantasy”) as an unfolding dynamic and complex activity of *imagining*. Imagining, according to Husserl, is not an awareness of internal representations, but the active engagement with objects, events and environments in the mode of simulation (“quasi-perception”). Imagining, like perceiving, is an intuitive (“anschauliche”) or sensory (“sinnliche”) experience – only, it engages the world not in the mode of actuality, but in the mode of possibility (Husserl 2005; Jansen 2015). As simulation of a possible experience, imagination may be just as embodied as perceptual cognition. What’s more, simulated “quasi-perception”, involves not only an intentional directedness towards, but also an embodied engagement with objects that are taken by the imagining subject (or, more functionally speaking, “imagining system”) to be unreal, absent, or possible, but nonetheless “external”, i. e., different

from the system itself and, while not located in the actually present perceptual field, nonetheless situated in a simulated environment. Furthermore, imagination, better: *imagining*, is a temporally extended activity that cannot occur in isolation from other activities. Although we may be more or less aware of it, whenever we are imagining we are also continuing to perceive. In fact, Husserl claims that some – however minimal – awareness of the difference between my imagining and my perceiving, and between the simulated and the actual environment, is even constitutive of genuine experiences of imagining (as opposed to experiences of dreaming or hallucinating). Typical “snapshot” analyses of imagination obscure these complexities and encourage the idea that imagination consists solely in taking oneself “out” of one’s body, or out of the perceptually present actual context or environment. However, even though my imagining might be taking me out of “my” body and into a body that can fly, or a body that is ten feet tall, etc., imagining, in the Husserlian sense of phantasy, cannot get me out of embodiment in general; neither can it take me out of any environment in general.<sup>4</sup> It is also for this reason that imagining, as simulation of possibly experience, is not the same as conceiving a possibility in thought.

Much more can and must be said about imagining, of course. And our understanding would be significantly enriched if we were to combine this phenomenological model with current imagination research in philosophy of mind. However, even this quick and cursory look at Husserl’s account alone, shows us that imagination might fit the 4E model much better than we might at first believe. What could that mean? It could mean that imagining is

embedded in simulated and, possibly, also actual environments; enacted as a, possibly, skilful activity (rather than a mere state); embodied, i. e., in part constituted (enabled and constrained) by features of embodiment; and, finally, extending not only beyond the brain, but often beyond the limits of the body and into the (possibly and/or actual) environment.

I take it that the last claim, viz. imagining may be extended, is the least immediately plausible of the four. Let’s therefore consider it.

Here the “enactive” view of imagining offered by Husserl may ease concerns we might have regarding phenomenal externalism: While it may be difficult to see how imaginative mental *states* could occur “outside” the head, or could even constitutively rely on events or processes “outside” the head; it might be somewhat easier to see how the *activity* of imagining could be extended beyond the “head”, that is, how it might not only use “external” objects and structures,

---

<sup>4</sup> In fact, as Sartre (2005) emphasizes, imagining always occurs in a relation to a reality I must be able to deny.



but also be partially constituted by them. As Andy Clark has recently pointed out in a rarely cited article (Clark 2005), imagining may be described as “extended” in exceptional cases, in which it involves and requires, for example, models. In these cases, such models do not function as external aids for imagining, but play a constitutive role in imagining. Clark is here thinking exclusively of “real world” models in the actual perceptual environment of the imagining system. However, if we let go of the idea that imagined objects and environments are internal to the mind, and if we instead think of imagining as an activity that simulates the engagement with external objects and environment, then we might be able to lift the “perception bias” that limits Clark’s important insight. Imagined objects and scenarios themselves may serve in a constitutive role for further imaginings. Put differently, simulated environments may ease the mental burden of having to represent all elements of an imagined object internally.

Furthermore, we may also think of these processes as intersubjectively extended. Imagined objects and simulated environments are not private mental items. In fact, we are all familiar with cases in which we share them. Think of fictions, for example. After all, it makes little sense to say that sleeping beauty is a private mental image in my head. From a Husserlian perspective, we can account for intersubjective sharing of imagined objects by means of a notion of the ideality of sense in general, which Husserl also uses against the psychologism of his day. According to this model, the sense of objects that are not, like public perceptual objects, individuated in space and time, is not therefore reducible to mental events. We not only co-perceive spatio-temporary objects, we are also able to co-intend the number “2”, whose sense is neither physically (publically, objectively) nor psychologically (privately, subjectively) real, but *ideal* and thus *intersubjectively* sharable. The more interesting question is therefore not whether and how it is possible to share *imagined objects*, but whether and how it is possible to share *in the imagining* of those objects. Again, I believe we all know examples that suggest that imagining may be extended intersubjectively even in this respect. Consider what we call “brain storming”, which we may do individually, but which more often than not is a collective event. It may also help here to think of imagining in more obviously embodied ways. Consider the activity of collectively and creatively imagining a piece or style of music, let’s say in shared improvisation; or consider the emergence of a dance in the inter-action and inter-responsiveness of dancing.

## 4 Conclusion

I have tried to make explicit some of the historical and current biases that have constrained imagination research in general and the inclusion of the imagination in the 4E model in particular. I have also pointed to as of yet largely untapped resources in phenomenology as well as current philosophy of mind that could help ease those constraints. However, perhaps the biggest stumbling block to accepting the idea of 4E imagination might be our reluctance to give up a tacit investment in the idea that imagination is a specifically human creative power that is immune not only to functionalist accounts, but also to extensions beyond the individual “free” mind of the imagining subject. While I believe that in important respects it remains the case, even if we use a 4E model, that I am the free author of my imaginings, I also believe that it is important to acknowledge that this view requires some qualification. Imagining is, at least partly, a distributed process which may involve extra-mental factors as well as social institutions, such as language, culture, ideologies, scientific paradigms, etc. (Gallagher 2011). This means that what enables and constrains my supposedly “free” imagination is not merely myself, not even only my embodied self. It also includes the natural and cultural landscapes in which I am embedded, and in which my imagining is embedded, as well as the activities I am able to perform, and the various “extensions” I have at my disposal. In other words, what is imaginable and what is not imaginable at any given time, has to do with far more than with the imaginer’s subjective abilities, which ultimately is not only an academic or scientific issue, but also a political one.

## Bibliography

- Abela, Paul (2002): *Kant’s Empirical Realism*. Oxford: Oxford University Press.
- Adorno, T.W. (1982): *Against Epistemology: A Metacritique*. Oxford: Blackwell.
- Allais, Lucy (2015): *Manifest Reality: Kant’s Idealism and his Realism*. Oxford: Oxford University Press.
- Allison, Henry (2004): *Kant’s Transcendental Idealism*, revised and enlarged edition. New Haven and London: Yale University Press.
- Banham, Gary (2006): *Kant’s Transcendental Imagination*. Basingstoke: Palgrave Macmillan.
- Bennett, Jonathan (1966): *Kant’s Analytic*. Cambridge: Cambridge University Press.
- Bernet, Rudolf (2004): “Husserl’s Transcendental Idealism Revisited”. In: *New Yearbook for Phenomenology and Phenomenological Philosophy* 4, pp. 1–20.
- Bird, Graham (2006): *The Revolutionary Kant*. Peru, Ill.: Open Court.
- Clark, Andy (2005): “Beyond the Flesh: Some Lessons from a Mole Cricket”. In: *Artificial Life* 11 (1–2), pp. 233–244.

- Clark, Andy and Torribio, Josefa (1995): “Doing Without Representing?” In: *Synthese*, vol. 101, pp. 401–431.
- Dummett, Michael (1993): *The Origins of Analytical Philosophy*. Cambridge, MA: Harvard University Press.
- Føllesdal, Dagfinn (1969): “Husserl’s Notion of Noema”. In: *Journal of Philosophy*, 66 (20), pp. 680–687.
- Gallagher, Shaun (2011): “The Overextended Mind”. In: *Versus* 113, pp. 57–68.
- Gallagher, Shaun (2009): “Philosophical Antecedents to Situated Cognition”. In: *Cambridge Handbook of Situated Cognition*, ed. by Philip Robbins and Murat Aydede. Cambridge: Cambridge University Press, pp. 35–51.
- Gibbons, Sarah L. (1994): *Kant’s Theory of Imagination*. Oxford: Clarendon Press.
- Hanna, Robert (2006): *Kant, Science and Human Nature*. Oxford: Oxford University Press.
- Heidegger, Martin (2010): *Kant und das Problem der Metaphysik*, ed. by Friedrich-Wilhelm von Herrmann. Frankfurt a. M.: Klostermann.
- Husserl, Edmund (2005). *Phantasy, Image Consciousness, and Memory*, trans. and ed. by John Brough. Dordrecht: Springer.
- Husserl, Edmund (1980): *Phantasie, Bildbewusstsein, Erinnerung*. Zur Phänomenologie der anschaulichen Vergegenwärtigungen. Texte aus dem Nachlass (1898–1925), ed. by Eduard Marbach. The Hague: Martinus Nijhoff.
- Husserl, Edmund (1973): *Die Idee der Phänomenologie*. Fünf Vorlesungen, ed. by Walter Biemel. The Hague: Martinus Nijhoff.
- Husserl, Edmund (1970): *Crisis of the European Sciences and Transcendental Phenomenology*, trans. By David Carr. Evanston: Northwestern.
- Husserl, Edmund (1956): *Erste Philosophie (1923/4)*. Erster Teil: Kritische Ideengeschichte. The Hague: Martinus Nijhoff.
- Husserl, Edmund (1950): *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*. Erstes Buch: Allgemeine Einführung in die reine Phänomenologie, ed. by Walter Biemel. The Hague: Martinus Nijhoff.
- Jansen, Julia (2015): “Imagination – Phenomenological Approaches”. In: *Routledge Encyclopedia of Philosophy*.  
<https://www.rep.routledge.com/articles/imagination-phenomenological-approaches>, visited on 25 May 2015.
- Jansen, Julia (2013): “Imagination, Embodiment and Situatedness: Using Husserl to Dispel (Some) Notions of “Off-Line Thinking””. In: *The Phenomenology of Embodied Subjectivity*, ed. by Rasmus T. Jensen and Dermot Moran. Dordrecht: Springer.
- Kind, Amy (2016): *Routledge Handbook of the Philosophy of Imagination*. London and New York: Routledge.
- Kind, Amy/Kung, Peter (2016): *Knowledge Through Imagination*. Oxford: Oxford University Press.
- Körner, Stephan (1987): “On Brentano’s Objections to Kant’s Theory of Knowledge”. In: *Topoi* 6, pp. 11–17.
- Kraft, Julius (1957, 2<sup>nd</sup> ed.): *Von Husserl zu Heidegger: Kritik der phänomenologischen Philosophie*. Frankfurt a. M.: Verlag Öffentliches Leben.
- Langton, Rae (1998): *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford: Oxford University Press.

- Longuenesse, Beatrice (1998): *Kant and the Capacity to Judge*. Princeton: Princeton University Press.
- Luft, Sebastian (2011). *Subjectivity and Lifeworld in Transcendental Phenomenology*. Evanston: Northwestern University Press.
- Makkreel, Rudolf A. (1990): *Imagination and Interpretation in Kant*. Chicago: Chicago University Press.
- Merleau-Ponty, Maurice (1962): *Phenomenology of Perception*, Colin Smith (trans.). New York: Humanities Press.
- Moore, G.E. (1903): "The Refutation of Idealism". In: *Mind* 12, pp. 433–53.
- Moran, Dermot (2008): "Husserl's Transcendental Philosophy and the Critique of Naturalism". In: *Continental Philosophy Review* 41, pp. 401–425.
- Naberhaus, Thane M. (2008): "Husserl's Transcendental Idealism". In: *Husserl Studies* 23, pp. 247–260.
- Rowlands, Mark (2010): *The New Science of the Mind*. Cambridge, MA: MIT Press.
- Sartre, Jean-Paul (2004): *The Imaginary*, trans. by Jonathan Webber. London and New York: Routledge.
- Schulking, Denis/Verburgt, Jacco (Eds.) (2011): *Kant's Idealism. New Interpretations of a Controversial Doctrine*. Dordrecht: Springer.
- Shapiro, Lawrence (2011): *Embodied Cognition*. London and New York: Routledge.
- Smith, David W./McIntyre, Ronald (1982): *Husserl and Intentionality: A Study of Mind, Meaning, and Language*. Dordrecht and Boston: Reidel.
- Strawson, Peter F. (1966): *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. London and New York: Routledge.
- Welton, Donn (Ed.) (2003): *The New Husserl: A Critical Reader*. Bloomington and Indianapolis: Indiana University Press.
- Welton, Donn (1998): *The Other Husserl*. Bloomington and Indianapolis: Indiana University Press.
- Wilson, Margaret (2002): "Six views of Embodied Cognition". In: *Psychonomic Bulletin & Review* 9 (4), pp. 625–636.
- Zahavi, Dan (2002): *Husserl's Phenomenology*. Stanford: Stanford University Press.



Thorsten Streubel

# Intuition und Argumentation – zum Verhältnis von intuitiver und diskursiver Vernunft

**Abstract:** *Intuition and argumentation – on the relation between intuitive and discursive reason.* In contemporary philosophical discussions participants often refer to their “intuitions” related to a special topic. Amazingly, “intuition” is nowadays by no means a clear epistemic or methodological concept, although there is a long tradition (from Plato to Husserl and Scheler), that understands intuition as a basic kind of reason. In this paper I want to rehabilitate “intuition” – through a historic anamnesis (Descartes, Kant, Husserl) – as an epistemic-methodological concept.

In kaum einer anderen philosophischen Richtung wurde und wird so häufig von Intuitionen gesprochen wie in der Analytischen Philosophie. Man würde daher mit Fug annehmen, dass es sich um ein klar definiertes epistemisches und/oder methodisches Konzept handeln würde. Dem ist aber erstaunlicherweise nicht so. Anders als etwa in der Phänomenologie (in Form von Husserls Methode der eidetischen Variation) gibt es in der Analytischen Philosophie keine vergleichbar ausgearbeitete Methodologie und Epistemologie der Intuition.<sup>1</sup> Allerdings wird der Terminus ‚Intuition‘ in analytisch geprägten Diskussionen auch nicht völlig beliebig gebraucht, sondern zumeist in dem Sinne, dass es sich bei Intuitionen um Überzeugungen handelt, die irgendwie aus direkter Sacheinsicht herrühren, in dessen in der Regel unmethodisch gewonnen und noch nicht hinreichend gerechtfertigt wurden. Anders verhält es sich hinsichtlich der Theoretisierung dessen, was mit Intuition gemeint ist, in der alteuropäischen Nous-Tradition von Platon/Aristoteles bis Husserl und Scheler. Intuition (noein) meinte hier so etwas wie geistige Einsicht in das, *was ist*. So konnten die obersten Seinsprinzipien wie etwa das Widerspruchsprinzip (ausgedrückt im Satz vom Widerspruch<sup>2</sup>) oder mathematische Axiome, aber auch die Erfassung des Eidos einer Sache nicht auf rein diskursivem Wege gewonnen werden, sondern nur unter Zuhilfenahme der noetischen Intuition. Intuition wird hierbei nicht im Sinne von spontanen Ein-

---

<sup>1</sup> Zur analytischen Diskussion des Intuitionsbegriffs vgl. z.B. Cappelen (2012) und Chudnoff (2013).

<sup>2</sup> Vgl. Met. 1005b ff.

fällen und Eingebungen verstanden, sondern als höchste Form der Sacheinsicht. Die Fähigkeit zur Intuition oder zur intellektuellen Sachverhalts-Einsicht wurde dabei als ein Monopol des Menschen verstanden, die ihm kraft seiner Geistnatur (Nous-Habe) zukam. Noch Scheler wird hierin die Sondernatur des Menschen erblicken. Zugleich gab es aber auch immer wieder Bemühungen die Intuition methodisch zu disziplinieren und einen methodologischen Begriff der Intuition zu erarbeiten. Ansätze hierzu finden sich bereits bei Platon und Aristoteles, besonders aber bei Descartes (in den *Regulae*), bei Kant (mathematische Methodik) und natürlich bei Husserl. Die Philosophie macht sich hiernach ein Grundvermögen des Menschen methodisch zunutze, um zu echter, aus „den Sachen selbst“ geschöpfter, Erkenntnis zu gelangen.

Um den durchaus rationalen Aspekt des Intuitionsbegriffs herauszustreichen, möchte ich im Folgenden zwei Vernunftbegriffe unterscheiden: Intuitive und diskursive Vernunft. Mit intuitiver Vernunft meine ich dabei eben dasjenige Vermögen, welches die Tradition als Nous oder Intellekt bezeichnet hatte, also eine Art unmittelbarer Einsichtsfähigkeit. Die intuitive Vernunft wäre damit die angeborene Disposition zu möglichen Intuitionen als noetischen Leistungsvollzügen. Unter ‚diskursiver Vernunft‘ möchte ich dagegen die Fähigkeit verstehen, Sätze oder Urteile in einen logischen Begründungszusammenhang zu bringen, also im logischen Sinne zu argumentieren und zu schließen, aber auch die hierbei vorausgesetzte Fähigkeit, zu urteilen.<sup>3</sup> – Allerdings ist gleich vorweg darauf hinzuweisen, dass die diskursive Vernunft konstitutiv auf die intuitive Vernunft bei ihren Vollzügen angewiesen ist und daher nicht einfach in einen Gegensatz zur Intuition gebracht werden darf.

Mir geht es nun im Folgenden um *vorbereitende* Überlegungen zur Gewinnung eines scharf umrissenen und methodisch brauchbaren Intuitionsbegriffs. Im Rahmen einer *ausgearbeiteten* Theorie der Intuition gälte es dann mindestens die folgenden drei Aufgaben zu bewältigen:

Es wäre *erstens* zu zeigen, dass die intuitive Vernunft in einem zweifachen Sinne konstitutiv für alle Formen der Diskursivität ist:

- a) Ohne intuitive Vernunft gäbe es auch keine diskursive Vernunft, da schon der Spracherwerb sowie einfache Begriffsbildungen das Einsichtsvermögen voraussetzen.
- b) Auch das aktuelle Verstehen von sprachlichen Äußerungen sowie die Erfassung logischer Zusammenhänge kann als Form der intuitiven Erfassung

---

<sup>3</sup> So besteht die phänomenologische Deskription darin, Urteile aufgrund von anschaulichen Gegebenheiten und mit Hilfe der Intuition zu fällen, ohne dass die einzelnen Urteile in einem deduktiven Zusammenhang stünden.

verstanden werden, so dass Diskursivität ohne Intuition (als hermeneutische Fähigkeit des unmittelbaren Verstehens und Erfassens) auch synchron unmöglich wäre.

*Zweitens* wäre zu zeigen, dass echte Begründungen letztlich nicht immer wieder auf weitere Sätze rekurrieren können, da dies einen unendlichen Regress implizieren würde. Vielmehr sind es die (freilich revisionsfähigen) anschauungsfundierten Sätze, in denen alle diskursiven Begründungen terminieren müssen. Diese aber sind ohne Anschauung und intuitive Einsicht nicht zu gewinnen.

*Drittens* gälte es, die Intuition als methodisches Konzept innerhalb der Philosophie wieder fruchtbar zu machen. Der unreflektierte und undisziplinierte Gebrauch von Intuitionen (oder auch die bloßen Appelle an die Intuition) sollte durch eine methodische Praxis der Intuition ersetzt und diese daher zu einem grundlegenden methodischen Instrument entwickelt werden.

Aufgrund des knappen Raumes werde ich hier versuchen, anhand dreier historischer Beispiele zu zeigen, welche Leistungen die Intuition vollbringen soll und kann. Die historische Anamnese dient dabei einzig und allein dazu, den Intuitionsbegriff von seinem irrationalen Beigeschmack zu befreien und ihn als brauchbares methodisches Konzept zu rehabilitieren. Eine umfassende Theorie der Intuition kann ich hier freilich nicht liefern. Ich werde im Folgenden vielmehr Schlaglichter auf drei prominente Autoren der abendländischen Philosophiegeschichte werfen, bei denen sich methodologische Überlegungen zur Intuition finden: nämlich Kant, Descartes und Husserl.

Dass ich mich auch auf Kant beziehe, mag insofern erstaunlich anmuten, als gerade Kant das Konzept einer intuitiven Vernunft als eines anthropologischen Vermögens strikt abgelehnt hat. Die menschliche Vernunft ist nach Kant ein rein diskursives Vermögen, welches sowohl in theoretischen wie in praktischen Zusammenhängen die Leitung in den menschlichen Angelegenheiten besitzen sollte. Er identifiziert es unter anderen mit folgenden Fähigkeiten: Urteilen, Schließen, Berechnen, Gründe geben und fordern, Gesetzgebung im Bereich der Natur und des Handelns, Abwägen von Handlungsalternativen, Maximenbildung, Entschlussfassung etc. Die Vernunft als ein *diskursives* Vermögen zu verstehen bedeutet hiernach, sie als die Fähigkeit zu verstehen, propositionale Gehalte zu erzeugen, die sie dann aufeinander zu beziehen und dadurch neue Propositionen zu generieren vermag, die entweder selbst das Ziel ihrer Tätigkeit bilden oder handlungsermöglichende und handlungsdirigierende Funktion haben. Je nachdem welches dieser Ziele man verfolgt, bedient man sich der Vernunft in theoretischer oder praktischer Absicht. Und aus diesen unterschiedlichen Verwendungen resultiert die funktionale Unterscheidung zwischen theoretischer und praktischer Vernunft. Sowohl theoretische als auch praktische Vernunft verfahren



diskursiv. Und doch spielt die Vernunft als intuitives Vermögen der Sache nach bei Kant eine zentrale Rolle – nicht nur in der Mathematik, sondern auch im Bereich der empirischen Erfahrung. Hier tritt sie verdeckt unter dem Namen der Urteilskraft auf (worauf ich hier aber nicht näher eingehen kann). Ich werde mich im Folgenden auf die Rolle der Vernunft im Rahmen von Kants Philosophie der Geometrie konzentrieren.

## 1 Kant

Kant behauptet in der *Kritik der reinen Vernunft* apodiktisch: „Der Verstand vermag nichts anzuschauen, und die Sinne nichts zu denken. Nur daraus, dass sie sich vereinigen, kann Erkenntnis entspringen.“ (KrV B 75) Dieses zentrale Lehrstück Kants, so überzeugend es auf den ersten Blick sein mag, ist bei genauerem Hinsehen jedoch höchst problematisch, ja aus meiner Sicht unhaltbar, zumal Kant selbst diese rigorose Funktionenaufteilung insbesondere in seiner Theorie der Mathematik relativiert. Nicht dass die Sinne in der Mathematik anfangen zu denken, aber die Vernunft ist hier plötzlich fähig, anzuschauen – oder sagen wir besser: Einsicht zu nehmen in mathematische Sachverhalte.

Die Mathematik ist zwar nach Kant „ein reines Produkt der Vernunft“ und „durch und durch synthetisch“ (Prol. § 6). Im § 7 der *Prolegomena* heißt es jedoch:

Wir finden aber, dass alle mathematische Erkenntnis dieses Eigentümliche habe, dass sie ihren Begriff vorher *in der Anschauung* und zwar *a priori*, mithin einer solchen, die nicht empirisch, sondern reine Anschauung ist, darstellen müsse, ohne welches Mittel sie nicht einen einzigen Schritt tun kann; daher ihre Urteile jederzeit *intuitiv* sind, anstatt dass Philosophie sich mit *diskursiven* Urteilen *aus bloßen Begriffen* begnügen und ihre apodiktischen Lehren wohl durch Anschauung erläutern, niemals aber daher ableiten kann. Diese Beobachtung in Ansehung der Natur der Mathematik gibt uns nun schon eine Leitung auf die erste und oberste Bedingung ihrer Möglichkeit; nämlich es muss ihr *irgendeine reine Anschauung* zum Grunde liegen, in welcher sie alle ihre Begriffe *in concreto* und dennoch *a priori* darstellen oder, wie man es nennt, sie *konstruieren* kann.

Was Kant damit meint, erläutert er ausführlich im ersten Hauptstück der *Transzendentalen Methodenlehre*<sup>4</sup> der *Kritik der reinen Vernunft* (der *Disziplin der reinen Vernunft*), wobei sich die Sachlage für Geometrie, Arithmetik und Algebra jeweils unterschiedlich darstellt. (Ich werde mich hier, wie gesagt, auf Kants Theorie der geometrischen Erkenntnis beschränken).

---

<sup>4</sup> „Ich verstehe also unter der transzendentalen Methodenlehre die Bestimmung der formalen Bedingungen eines vollständigen Systems der reinen Vernunft.“ (KrV B 736)

Die reine Mathematik wird dabei von Kant methodisch sowohl von der Metaphysik als auch der Physik abgegrenzt. Im Unterschied zur Metaphysik und ähnlich wie die Physik kann und muss die Mathematik auf die Anschauung rekurrieren. Die mathematischen Sätze werden aber – anders als die Hypothesen der Physik – nach Kant nicht erst nachträglich durch das Experiment bewährt oder falsifiziert, sondern die Mathematik ist methodisch fast durchweg anschauungsgebunden. In der Mathematik wird die Vernunft plötzlich selbst *anschauend* (einsehend) – und bleibt doch auf die (sinnliche) *Anschauung* als ihr funktionales Komplement angewiesen. Und das ist in Bezug auf Kants sonstige Auslassungen zum Thema Vernunft äußerst bemerkenswert. In gewisser Weise wird die menschliche Vernunft nämlich in der Mathematik zu einer Art *intuitus* originarius, also zu einer Vernunft, die ihre Objekte im Anschauen selbst erzeugt – oder präziser: die im Erzeugen der Objekte mittels Einbildungskraft diese Objekte anschaut und deren intelligiblen Strukturen erfasst. Nur Gott, falls es ihn gibt, verfügt eigentlich nach Kant über eine solche Vernunft, aber auf gar keinen Fall der Mensch. In seiner Philosophie der Mathematik feiert jedoch die von ihm selbst mit großem Aufwand und Aplomb eigenhändig gekreuzigte und beerdigte intellektuelle Anschauung ihre österliche Wiederauferstehung. Totgeglaubte leben länger, sagt der Volksmund. Und auch wenn es hier zunächst nur um das Fortleben eines Konzepts in einer prominenten philosophischen Gesamtkonzeption geht, so beweist doch Kants Theorie der Mathematik selbst augenfällig, dass wir über eine intuitive Vernunft oder den *Nous* verfügen, von dem Aristoteles überzeugt war, dass er das Göttliche im Menschen ist. In Kants Theorie der Mathematik wird die Existenz des *Nous* im Menschen de facto anerkannt, wenngleich ansonsten offiziell bestritten.

Inwiefern wird nun die Vernunft in der Mathematik nach Kant anschauend, inwiefern verfährt sie intuitiv?

Die *philosophische* Erkenntnis ist die *Vernunftkenntnis aus Begriffen*, die mathematische aus der *Konstruktion* der Begriffe. Einen Begriff aber konstruieren, heißt: die ihm korrespondierende Anschauung a priori darstellen. Zur Konstruktion eines Begriffs wird also eine *nicht empirische* Anschauung erfordert, die folglich, als Anschauung, ein *einzelnes* Objekt ist, aber nichts destoweniger, als die Konstruktion eines Begriffs (einer allgemeinen Vorstellung), Allgemeingültigkeit für alle mögliche Anschauungen, die unter denselben Begriff gehören, in der Vorstellung ausdrücken muss. (KrV B 741)

In Bezug auf die Geometrie heißt das nun Folgendes: Ausgangspunkt des geometrischen Beweises ist ein Begriff, etwa der des Dreiecks. Nach Kant haben die mathematischen Begriffe ihren Ursprung nicht selbst in der Anschauung oder gar in der empirischen Erfahrung, sondern als genuin *mathematische* Begriffe entspringen sie den jeweiligen Definitionen. Vor der Definition gibt es in der Ma-

thematik keine Begriffe. Mathematische Definitionen können daher nach Kant „auch niemals irren. Denn, weil der Begriff durch die Definition zuerst gegeben wird, so enthält er gerade nur das, was die Definition durch ihn gedacht haben will.“ (KrV B 759) (Dass dies so nicht stimmen kann, werden wir gleich noch sehen.) Kant definiert nun den Begriff des Dreiecks als „Begriff von einer Figur, die in drei geraden Linien eingeschlossen ist“ (KrV B 744).

Das spezifisch mathematische Beweisverfahren nennt Kant *Demonstration* (von *demonstrare*: zeigen, darlegen, beweisen), da es beweist, indem zugleich etwas aufgezeigt wird, etwa ein mathematischer Sachverhalt, der zu einem synthetischen Urteil a priori, dem sog. Mathemata, führt (oder ein solches zumindest beweisend bestätigt) und zugleich die Quelle seiner Wahrheit ist.<sup>5</sup>

Die geometrische Demonstration beginnt also mit der Konstruktion eines mathematischen Begriffs in der Anschauung, entweder durch Imagination oder mittels Zirkel und Lineal auf Tafel oder Papier. Dies ist der eigentümliche Schematismus der mathematischen Begriffe, der nicht mit dem Schematismus der reinen Verstandesbegriffe verwechselt werden darf. Der mathematische Begriff ist aber nach Kant, wie die meisten Begriffe, eine Erzeugungsregel für die Einbildungskraft. Und daher ist er, wie die meisten Begriffe, selbst eine Art Schema. Der Begriff ist die Erzeugungsregel seines Umfangs, also des Inbegriffs derjenigen *möglichen* Gegenstände, die unter ihn fallen. Das, was unter einen Begriff fällt, wird durch den Begriffsinhalt bestimmt. Der Begriff des Dreiecks ermöglicht dem Mathematiker nun, alle *möglichen* Dreiecke in der Anschauung zu erzeugen.

So konstruiere ich einen Triangel, indem ich den diesem Begriffe entsprechenden Gegenstand, entweder durch bloße Einbildung, in der reinen, oder nach derselben auch auf dem Papier, in der empirischen Anschauung, beidemal aber völlig a priori, ohne das Muster dazu aus irgendeiner Erfahrung geborgt zu haben, darstelle. Die einzelne hingezzeichnete Figur ist empirisch, und dient gleichwohl den Begriff, unbeschadet seiner Allgemeinheit, auszudrücken, weil bei dieser empirischen Anschauung immer nur auf die Handlung der Konstruktion des Begriffs, welchem viele Bestimmungen, z.E. der Größe, der Seiten und der Winkel, ganz gleichgültig sind, gesehen, und also von diesen Verschiedenheiten, die den Begriff des Triangels nicht verändern, abstrahiert wird. (KrV 741f.)

Auch wenn Kant nicht hinreichend deutlich zwischen einem empirischen Dreieck und einem singulären idealen geometrischen Dreieck unterscheidet, so ist doch klar, dass es nach ihm auf den idealen Gehalt der empirischen Figur ankommt und auf die Tatsache, dass auch ein ideales Dreieck die Vereinzelung des allgemeinen Begriffs „Dreieck“ darstellt. Der Gegenstand des geometrischen Begriffs ist je-

---

5 „Nur ein apodiktischer Beweis, so fern er intuitiv ist, kann Demonstration heißen.“ (KrV B 762)

denfalls nach Kant *nicht* die perzeptiv oder imaginativ vorstellige Figur – diese ist nur das „Schema“<sup>6</sup> eines idealen Dreiecks. Ein einzelnes ideales Dreieck ist dagegen weder eine empirische noch eine begriffliche Größe, wenngleich es selbst ein Moment des Begriffs(umfangs) „Dreieck“ ist. Es ist eine ideale Figur im Raum, die ihre anschauliche Präsenz freilich nur über ihr sinnliches Schema, das empirische Dreieck, gewinnt.

Im Unterschied zur Metaphysik nun, die das Besondere nur im Allgemeinen, also durch rein begriffliches Denken, erfassen kann, *betrachtet* die Mathematik „das Allgemeine im Besonderen, ja gar im Einzelnen, gleichwohl doch a priori und *vermittelt* der Vernunft“ (KrV B 742; herv. T.S.).

Kant illustriert seine These von der Anschauungsfundiertheit der Mathematik am Beispiel des Beweises von der Innenwinkelsumme euklidischer nichtsphärischer Dreiecke. Den durch Definition gebildeten Begriff kann man freilich analysieren und das zum Vorschein bringen, was man zuvor in ihn investiert hat. Aber man erhält hierdurch keine valide Auskunft über den gesuchten geometrischen Sachverhalt. Begriffszergliederung führt zu analytischen, nicht aber zu synthetischen, also erkenntniserweiternden Urteilen. An dieser Stelle wird freilich deutlich, dass Kants These von der Infallibilität mathematischer Begriffe schlicht falsch ist: Denn wenn ich einen Begriff des Dreiecks bilde, bei dem die Innenwinkelsumme größer oder kleiner als  $180^\circ$  beträgt, dann habe ich einen unrichtigen Begriff gebildet. Auch die Möglichkeit mathematischer Gegenstände – als Korrelate mathematischer Begriffe – muss daher die Anschauung verbürgen. Dreiecke haben eben eine Eigennatur, die nicht von Menschenvernunft gemacht oder willkürlich geändert werden kann.

Um zu zeigen, wie groß die Innenwinkelsumme bei allen euklidischen nichtsphärischen Dreiecken ist, wird der Geometer also ein einzelnes Dreieck konstruieren.

Weil er weiß, dass zwei rechte Winkel zusammen gerade so viel austragen, als alle berührende Winkel, die aus einem Punkte auf einer geraden Linie gezogen werden können, zusammen, so verlängert er eine Seite seines Triangels, und der bekommt zwei berührende Winkel, die zweien rechten zusammen gleich sind. Nun teilet er den äußeren von diesen Winkeln, indem er eine Linie mit der gegenüberstehenden Seite des Triangels parallel zieht, und sieht, dass hier ein äußerer berührender Winkel entspringe, der einem inneren gleich ist usw. Er gelangt auf solche Weise durch eine Kette von Schlüssen, immer von der Anschauung geleitet, zur völlig einleuchtenden und zugleich allgemeinen Auflösung der Frage. (KrV B 744 f.)

---

**6** Der Schemabegriff wird bei Kant nicht immer einheitlich gebraucht: Schema kann einerseits ein Begriff sein, insofern er die Erzeugungsregel seines Umfangs darstellt. Andererseits aber bezeichnet Kant mit Schema schematische sinnliche Gestalten.

Dieser Beweis wird zwar nur *an einem* einzelnen Dreieck geführt, aber er wird dadurch zugleich *für alle* Dreiecke geführt. Und diese *reine* Induktion kann nur die ideative intuitive Vernunft leisten: Am Ende des Beweises muss man *einsehend* erfassen, dass eine Variation der Winkel und Seitenlängen des Dreiecks niemals den Beweis und seine Beweiskraft selbst affiziert. Und jeder muss es selbst einsehen. Die Einsicht wird demonstrativ ermöglicht und kann durch didaktisches Geschick erleichtert werden, aber „Sehen“ muss jeder für sich. Niemandem kann der Akt der Einsicht abgenommen werden. Aber nicht erst der gesuchte Sachverhalt muss einsehend erfasst werden, sondern die ganze Demonstration muss intuitiv-diskursiv nachvollzogen werden.

Man sieht, bei der geometrischen Demonstration sind mehrere Erkenntnisfunktionen involviert: Die Anschauung, die Einbildungskraft, Vorwissen, diskursive Vernunft – und eben die Fähigkeit, Sachverhalte in der Anschauung intuitiv (ein-sichtig) zu erfassen. Man muss hierbei nicht nur sinnlich sehen, sondern auch „geistig“, eben mit der intuitiven Vernunft. Man muss sehend-erfassend Ein-Sicht nehmen können in das, was (gegeben) ist. Anschauung und diskursive Vernunft wären zwei unverbunden sich drehende Räder, wenn sie nicht über ein Drittes immer schon verbunden wären. Und diese Verbindung leistet einzig und allein die intuitive Vernunft, ohne die jeglicher Schematismus ins Leere liefe, da durch Schematisierung zwar der Begriffsumfang erzeugt werden kann, aber dadurch allein noch lange nichts erkannt wird. Im Erkenntnisprozess müssen Anschauung, intuitive und diskursive Vernunft produktiv zusammenwirken, um Einsichten hervorzubringen. Aber die Kärnerarbeit hat die intuitive Vernunft zu verrichten. Sie sorgt für die Bodenhaftung. Sie ist der Blick ‚nach unten‘, auf die „Sachen selbst“. Jegliches Erschautes muss jedoch auch in einem diskursiven Urteil terminieren, wenn die Erkenntnis zum bleibenden Besitz werden und nicht auf halbem Wege stecken bleiben soll. Daher wäre Intuition ohne Diskursivität zwar weder blind noch leer, aber doch stumm im wörtlichen Sinne: sprachlos.

Bleibt noch zu ergänzen, dass nach Kant auch die Axiome der euklidischen Geometrie nicht selbstevident sind, sondern aus der Anschauung geschöpft werden. Die mathematischen Axiome („z. B. zwischen zwei Punkten kann nur eine gerade Linie sein“) werden „nur in der reinen Anschauung“ erkannt (KrV B 356 f.). Sie sind nicht in sich wahr, sondern drücken räumliche Grundsachverhalte aus, die nur durch Rückgang auf die Anschauung erfasst werden können. Das unterscheidet die euklidische Geometrie von allen nichteuklidischen Geometrien: Nur ihre Axiome und synthetischen Urteile a priori sind anschaulich fundiert und drücken nachvollziehbar bestehende Sachverhalte aus, während alle nichteuklidischen Geometrien zunächst einmal reine Konstruktionen sind, welche durch Variation der euklidischen Axiome generiert werden und lediglich durch ihre Nützlichkeit im Rahmen naturwissenschaftlicher Theoriebildungen gerechtfertigt

sind, aber deshalb nicht unbedingt ontologische Sachverhalte beschreiben. Man sollte sich daher vor deren vorschnellen Ontologisierung hüten (etwa: ‚Der wirkliche Raum *ist* nichteuklidisch‘). Der einzige Raum, der wirklich im strengen Sinne originär erfahrbar ist, ist der Anschauungsraum. Und in diesem allein konstruiert die euklidische Geometrie ihre Begriffe und gelangen auch nichteuklidische Räume zur Darstellung.<sup>7</sup>

Was lernen wir aus Kants Theorie der Geometrie für das Verhältnis von intuitiver und diskursiver Vernunft? Auch wenn beide Funktionen miteinander verflochten sind, so wäre doch eine rein diskursive Vernunft völlig blind (und nicht nur leer), wenn sie nicht über eine anschauende-einsehende Vernunft mit der Anschauung verbunden wäre. Die Anschauung mag ohne Begriffe blind sein. Aber ohne intuitive Vernunft wäre eben auch die diskursive Vernunft völlig blind. Erst die intuitive Vernunft lässt die diskursive Vernunft ‚sehen‘ und dann über das Eingesehene urteilen. In unserem Beispiel wusste der Geometer zwar von Anfang an, wonach er zu suchen hatte, aber dass die Innenwinkelsumme genau  $180^\circ$  beträgt, wurde weder diskursiv bewiesen noch empirisch gemessen, sondern musste anhand der anschaulichen Konstruktion eingesehen werden.

## 2 Descartes

Anders als Kant versteht Descartes die Intuition noch als grundlegende und nicht nur auf die Mathematik eingeschränkte Vernunftfunktion. Descartes geht von der Einheit des Intellekts aus, der sowohl intuitiv als auch diskursiv zu verfahren vermag. Und beide intellektiven Betätigungsformen gilt es methodisch zu disziplinieren und so miteinander zu verknüpfen, dass die Erlangung echten Wissens gewährleistet ist. In den *Regulae ad directionem ingenii* führt Descartes Intuition und Deduktion als die beiden Verstandeshandlungen an, aus deren Zusammenspiel gesichertes Wissen und damit Wissenschaft entspringen soll. Der Intuition kommt dabei der methodische Primat zu, denn die Deduktion soll nicht nur von

---

<sup>7</sup> Allerdings ist hierzu zu sagen, dass die Geometrie natürlich alles andere als eine Phänomenologie des Raumes darstellt. So wie sie sich nicht für empirische Dreiecke, sondern nur für ideale interessiert, so beschreibt sie insgesamt den idealen Raum, den man sich als entperspektivierten Anschauungsraum vorstellen muss. So wie ein ideales Dreieck nur in einem empirischen zur Darstellung kommen kann, so der euklidische Raum nur im Anschauungsraum, etwa als Koordinatensystem, das selbst nur einseitig-perspektivisch gezeichnet werden kann. Das beeinträchtigt aber nicht den Wahrheitsgehalt der euklidischen Geometrie, da der euklidische Raum immerhin perspektivisch zur Darstellung kommen kann, während nichteuklidische Räume nur euklidisch dargestellt werden können.

(durch Intuition gewonnenen) evidenten Einsichten ausgehen, sondern der deduktive Übergang, also das eigentliche Deduzieren, ist nach Descartes selbst nur durch Intuition in die jeweiligen logischen Verhältnisse möglich. Und hieraus wird schon ersichtlich, dass die Deduktion bei Descartes keinen autonomen Vollzug darstellt, sondern konstitutiv durch die Intuition realisiert wird. Während Descartes unter Deduktion „all das [fasst], was aus etwas anderem sicher Erkannten mit Notwendigkeit erschlossen wird.“ (R 3, 8), versteht er unter Intuition

nicht das schwankende Zeugnis der sinnlichen Wahrnehmung oder das trügerische Urteil der verkehrt verbindenden Einbildungskraft, sondern ein so müheloses und deutlich bestimmtes Begreifen des reinen und aufmerksamen Geistes, dass über das, was wir erkennen, gar kein Zweifel übrigbleibt, oder, was dasselbe ist: eines reinen und aufmerksamen Geistes unbezweifelbares Begreifen, welches allein dem Lichte der Vernunft entspringt [...]. So kann jeder intuitiv mit dem Verstande sehen, dass er existiert, dass er denkt, dass ein Dreieck von nur drei Linien, dass die Kugel von einer einzigen Oberfläche begrenzt ist und Ähnliches. (R 3, 5)

Außerdem erfasst die Intuition die Begriffe einfacher Naturen wie Figur, Ausdehnung, Dauer und deren (onto)logische Verhältnisse zueinander (wie etwa: jeder Bewegung kommt notwendig eine Dauer zu bzw. zum Begriff der Bewegung gehört der Begriff der Dauer).

Aus dem Bisherigen wird nun dreierlei deutlich:

(1) Der Intellekt ist primär ein Vermögen der Einsicht (Intuition), die auch funktional das einsichtige Deduzieren mitermöglicht. Mit dem Verstand kann man „ein-sehen“, d. h. etwas erfassen, was es unabhängig vom Einsehen und Erfassen des Verstandes gibt. Er ist somit kein *intuitus originarius* im Sinne Kants, aber eben auch kein rein diskursives Vermögen.

(2) Damit steht die Intuition (als Erkenntnisakt) für Descartes in einer funktionalen Mittelposition: Sie ist weder die bloße anschauliche (sei es perzeptive oder imaginative) Präsenz von etwas, noch verfährt sie nach Descartes selbst deduktiv, sondern eben intuitiv. Indem sie aber einfache Sachverhalte erfasst, kann sie Anschauung und diskursives Denken (bzw. die Deduktion) verknüpfen. Dies setzt freilich voraus, dass sie den Übergang zur Diskursivität dadurch bewerkstelligt, dass sie entweder das Fällen entsprechender prädikativer Erkenntnisurteile eigenständig vollzieht (und damit selbst diskursiv wird) oder dieses zumindest konstitutiv ermöglicht. Auch wenn Descartes auf dieses Problem nicht näher eingeht, so sind doch intuitive und diskursive Vernunft für ihn keine völlig getrennten Fakultäten, sondern Betätigungsformen des einen Geistes (*ingenium, intellectus*).

(3) Descartes identifiziert somit Intuition und Deduktion nicht miteinander oder erklärt die Deduktion zu einer besonderen Form der Intuition, aber er betont doch die funktionale Abhängigkeit des Schlussfolgerns und Deduzierens selbst von der Intuition:

[Die] Evidenz und Gewissheit der Intuition [wird] nicht allein bloß für Aussagen, sondern auch für jedes beliebige folgernde Denken erfordert. Denn wenn man z. B. diese Schlussfolgerung nimmt: zwei und zwei ergeben dasselbe wie drei und eins, so muss man nicht nur intuitiv sehen, dass zwei und zwei vier ergeben und dass drei und eins ebenfalls vier ergeben, sondern darüber hinaus, dass aus diesen beiden Propositionen jene dritte mit Notwendigkeit folgt. (R 3, 7)

Die Intuition stellt also nicht nur evidente Axiome, Prinzipien, Prämissen und klare und deutliche Begriffe für die Deduktion bereit, sondern spielt auch bei jedem schlussfolgernden Denken eine konstitutive Rolle. Und dies ist vor allem auch normativ gemeint: Descartes sucht ja nach einer verbindlichen wissenschaftlichen Methode. Je weniger Intuition im Schlussfolgern zum Einsatz kommt, desto höher wird die Fehlerwahrscheinlichkeit (die allerdings durch selbst einsichtig begründete Rechenmethoden wieder minimiert werden kann). Für Descartes ist es ein und derselbe Verstand (*intellectus*), der einsieht, urteilt und schlussfolgert. Und beim Urteilen und Schlussfolgern sollte die Intuition maßgeblich beteiligt sein, da diese diskursiven Akte sonst blind vollzogen würden und daher keine eigentlichen Erkenntnisvollzüge mehr darstellten.

Die Intuition ermöglicht also die Erfassung einfacher Naturen (wie Figur, Dauer etc.) und deren Verbindung in einfachen Sachverhalten (z. B. die notwendige Ausgedehntheit von Körpern), die hierauf gründende Bildung von einsichtigen (evidenten) Urteilen (Körper sind notwendig ausgedehnt), das Deduzieren neuer Erkenntnisurteile aus evidenten Urteilen und (in Verbindung mit der Fähigkeit zur Diskursivität) sogar das Verstehen von sprachlichen Äußerungen: Denn diese beiden Fähigkeiten (zur Intuition und Deduktion) sind bei aller methodischen Disziplinierung vorausgesetzt. Und könnte sich der Verstand ihrer „nicht schon vorher bedienen, [würde er] keine noch so einfache Vorschrift selbst der Methode verstehen“ (R 4, 2). Und letztlich müssen auch die Prinzipien der Logik ein Fundament in der Anschauung und der darauf bezogenen Intuition haben. Das heißt, nicht nur die logischen Beziehungen innerhalb bestimmter Urteilszusammenhänge müssen intuitiv erkannt werden, sondern etwa auch die apodiktische Wahrheit (und Reichweite) des Satzes vom Widerspruch, der Identität und vom ausgeschlossenen Dritten.

Die beiden entscheidenden Voraussetzungen des cartesischen Intuitionismus sind dabei, dass der Mensch über ein angeborenes „Licht der Vernunft“, also die Einsichtsfähigkeit (*intellectus*) als Disposition, verfügt (1); und dass die Erkenntnisgegenstände eine Eigennatur besitzen (2). Und nur weil letzteres der Fall ist, lässt sich auch herausfinden, ob gewisse begriffliche Zusammenhänge wesensnotwendig („Körper sind ausgedehnt“) oder nur empirisch notwendig („Körper sind schwer“) sind. Über wesensnotwendige Sachverhalte sagt Descartes, dass sie notwendig dann sind,



wenn der eine Sachverhalt gewissermaßen in einer solchen Verschlingung in den Begriff des anderen verwickelt ist, dass wir keinen von beiden deutlich vorstellen könnten, falls wir urteilen sollten, sie seien voneinander getrennt. Auf diese Weise ist die Figur mit der Ausdehnung verbunden, die Bewegung mit der Dauer oder der Zeit etc.; denn weder lässt sich eine Figur, der jede Ausdehnung fehlt, denken noch auch eine Bewegung ohne jede Dauer. Ebenso ist, wenn ich sage, 4 und 3 sind 7, diese Zusammensetzung notwendig, denn wir können die Anzahl Sieben nicht deutlich vorstellen, wenn wir nicht die Anzahlen Drei und Vier gewissermaßen verschlungen in sie einschließen. Und überhaupt steht alles, was über Gestalten und Zahlen bewiesen wird, auf diese Weise mit dem in notwendiger Verbindung, worüber man es aussagt. Auch findet sich diese Notwendigkeit nicht nur in den Gegenständen der sinnlichen Wahrnehmung, sondern auch, wenn z. B. Sokrates sagt, er zweifle an allem, so folgt hieraus notwendig: also erkennt er wenigstens dies, dass er zweifelt, und gleichfalls: also erkennt er, dass etwas wahr oder falsch sein kann usw., das ist nämlich mit der Natur des Zweifels notwendig verknüpft. (R 12, 17)

Das letzte Beispiel (Sokrates) zeigt wiederum, dass auch die diskursive Entfaltung von logischen Implikationen evidenter Einsichten bzw. Sätzen auf die intuitive Erfassung derselben angewiesen ist. Zu diesen Erkenntnissen kann natürlich nur ein kompetenter Sprecher kommen. Aber die Einsicht in die Wahrheit und Richtigkeit der so gewonnenen Erkenntnisse hängt nicht von der jeweiligen empirischen Sprache ab, sondern von der Natur der Sachen und den logischen Beziehungen der verwendeten Begriffe, die (also die logischen Beziehungen) eben nicht auf empirisch-grammatikalische Regeln relativiert werden oder gar mit diesen identifiziert werden dürfen. Aber natürlich gilt, dass auch die Erfassung grammatischer Strukturen nicht ohne die Intuition auskommt, da sie der unmittelbare Blick auf jegliche vorgegebene Gehalte ist – mögen diese nun bereits sprachlich-begrifflicher Natur sein oder nicht.

Nun spielt die Intuition bei Descartes zwar, wie wir gesehen haben, eine buchstäblich fundamentale Rolle im Rahmen seiner Methodologie, doch findet man bei ihm keine echte Phänomenologie und Epistemologie derselben. Insbesondere der Zusammenhang von einsichtiger Erfassung und dem Urteilen über das einsichtig Erfasste bleibt bei Descartes im Dunklen. Zudem findet man bei Descartes im Grunde auch keine ausgearbeitete Methodologie der Intuition als eines geistigen Vollzugs. Sie wird zwar im Rahmen der cartesischen Methodologie insofern methodisch diszipliniert, als sie auf die Erfassung der einfachen Naturen und deren Verbindungen sowie auf das einsichtige Schlussfolgern verpflichtet wird. Aber das Einsehen und die intuitive Erfassung selbst werden weder als weiter methodisch regulierbar verstanden noch in ihrer Natur aufgehellt. Mir scheint, dass erst im 20. Jahrhundert von Husserl der Versuch unternommen worden ist, eine umfassende Phänomenologie und Methodologie der Intuition als solcher zu geben.

### 3 Husserl: eidetische Variation und vorprädikative Erfahrung

Husserl thematisiert die Intuition als Erfassung des begrifflich Allgemeinen sowohl als vorphilosophische (alltägliche) Leistung (paradigmatisch: Wahrnehmung von etwas *als* etwas, aber auch Symbolverständnis) und als Grundvermögen des Menschen (Begriffsbildung) als auch als Möglichkeitsbedingung der Phänomenologie selbst. Die Phänomenologie (im husserlschen Sinne) zielt (rein formal betrachtet) auf die Gewinnung adäquater Beschreibungsbegriffe und die Feststellung notwendiger Zusammenhänge im Bereich des Phänomenalen ab. Daher gilt es, die Intuition als eingeborene Fähigkeit hinsichtlich ihrer Leistung phänomenologisch aufzuklären und methodisch fruchtbar zu machen. Dass hier freilich schwierigste Probleme vorliegen, dürfte schon allein daran sichtbar werden, dass bei einer Methodologie der Intuition die Gefahr eines Begründungszirkels insofern besteht, als die Phänomenologie sich durch ihre Methode allererst konstituiert, die sie jedoch selbst zuvor zu begründen hat. Ich kann hier aus Platzgründen auf diese methodologischen Schwierigkeiten leider nicht im Einzelnen eingehen, sondern nur darauf hinweisen, dass der Begründungszirkel dann vermieden werden kann, wenn man zwischen der Phänomenologie als methodisch begründeter Wissenschaft und der Phänomenologie als begründender Wissenschaft, d.i. als Methodologie, unterscheidet.<sup>8</sup> Stattdessen geht es mir hier lediglich darum, zu zeigen, dass die Intuition kein irrationales, dem rationalen Argumentieren entgegengesetztes, mysteriöses Schauen ist, sondern eine Form der Vernunft, ohne die auch die diskursive Vernunft undenkbar wäre.

Die ausführlichsten Analysen widmet Husserl einerseits dem Verhältnis von Wahrnehmung, vorprädikativer Erfahrung und prädikativer Erfahrung, andererseits der Methode der eidetischen Variation, die dazu dient, die reinen Begriffe der Phänomenologie zu konstituieren. Beide Bereiche gehören insofern zusammen, als auch die eidetische Variation als intuitives Verfahren in Prädikationen terminieren muss.

Ein wesentliches Ziel der eidetischen Variation ist es, die eidetischen Beziehungen und Verflechtungen dieser Begriffe zu erfassen. Oder gegenständlich ausgedrückt: Wie Descartes fragt Husserl, inwiefern bestimmte phänomenale Verhältnisse empirisch kontingent (bzw. hypothetisch notwendig) und inwiefern sie eidetisch-notwendig sind. Mit Letzterem ist gemeint: dass etwas gar nicht anders sein kann, dass sein Gegenteil unmöglich ist (in allen möglichen Welten).

---

<sup>8</sup> Vgl. hierzu detailliert Streubel (2016).

Dass Wahrnehmung immer Wahrnehmung von etwas ist, ist hiernach wesensnotwendig, weil eine Wahrnehmung von radikal gar nichts unmöglich wäre. Dass eine Wahrnehmung dagegen visuelle Inhalte aufweisen kann (oder eben nicht), ist empirisch kontingent, wie das Beispiel der Blindgeborenen beweist. Um zu validen Aussagen über notwendige Zusammenhänge zu gelangen, bedarf es auch nach Husserl der Intuition, von der er zeigt, dass ihre methodische Anwendung nicht in einem einfachen Hinblicken und Erfassen bestehen kann wie Descartes (und auch noch der Husserl der *Logischen Untersuchungen*) glaubte, sondern in einem systematischen Variieren von Teilmomenten der untersuchten Gegebenheiten. Um z. B. herauszufinden, was zur sinnlichen Wahrnehmung notwendiger Weise gehört, muss ich alle faktischen Momente variieren: die verschiedenen Arten der inhaltlichen Qualitäten, die Formen ihrer Gegebenheit (Raum, Zeit), ihre Klarheitsgrade, den Leib (d.i. dessen Struktur und kinästhetischen Funktionen) als Wahrnehmungsvoraussetzung, die begrifflichen Gehalte etc. Hier wird sofort deutlich, dass es nicht nur einen sehr allgemeinen Begriff von Wahrnehmung überhaupt zu bilden gilt, sondern auch spezifische Wahrnehmungsbegriffe. Es handelt sich hier um eine potentiell unendliche Aufgabe, wenngleich es durchaus möglich ist, einfache Notwendigkeitsbeziehungen zu konstatieren (etwa: kein Wahrgenommenes ohne Dauer). Die husserlsche Phänomenologie ist daher nicht nur eine analytische Wissenschaft des Wirklichen, sondern auch des prinzipiell Möglichen (und in beiden Bereiche gibt es notwendige, kontingente und unmögliche Zusammenhänge).

Intuitiv ist diese Methode nun deshalb, weil nicht einfach zweifelhafte Argumente angeführt werden, was aus irgendwelchen anfechtbaren Gründen der Fall sein muss, sondern weil die „Sachen selbst“ analysiert werden, ohne dass damit aber ein genereller Infallibilismus verbunden wäre. Doch ist es eben ein Unterschied, ob man sich die Dinge, über die man spricht, genau anschaut oder nicht. Die „Sachen selbst“ sind jedoch, wie gesagt, nicht nur die aktuell vorliegenden, sondern alle *möglichen*, soweit man sie sich prinzipiell anschaulich vorstellen kann. Daher dient etwa eine aktuelle Wahrnehmung (um bei diesem Beispiel zu bleiben) nur als Ausgangsexempel, dessen einzelne Momente umfingiert werden, um so die eidetischen Beziehungen der einzelnen Momente einsichtig zu erfassen. So kann ich etwa die *Inhalte* der aktuellen Wahrnehmung beliebig variieren, stets bleibt die Wahrnehmung eine Wahrnehmung, solange sie überhaupt irgendwelche Inhalte hat. Nur wenn ich jeglichen Inhalt versuche wegzufingieren, um die Vorstellung einer leeren Wahrnehmung zu erhalten, dann sehe ich sofort, dass mit dem Inhalt auch die Wahrnehmung verschwindet. Es ist unmöglich, sich eine inhaltslose Wahrnehmung vorzustellen – und dies liegt nicht an der Begrenztheit des Vorstellens, sondern weil die Natur der Wahrnehmung es unmöglich macht, sich etwas vorzustellen, was unmöglich ist. Diese *sachliche*

Unmöglichkeit wird selbst einsichtig erfasst im scheiternden Versuch, sie sich klar und deutlich als Möglichkeit vorzustellen. So ist es auch keine nur subjektive Unmöglichkeit sich unausgedehnte Körper (oder dauerlose Töne) vorzustellen, sondern unausgedehnte Körper sind keine Körper (sondern, wenn überhaupt, punktuelle Entitäten) – und dies wird evident, indem man sich unausgedehnte Körper vorzustellen versucht und erkennt, dass Ausdehnung etwas ist, was zur Natur der Körper und damit zu ihrem Wesensbegriff gehört.

Husserls methodologischer Grundgedanke in Bezug auf die Intuition ist also, dass das Erfassen einfacher Naturen und notwendiger Sachverhalte nicht selbst einfach ist, nicht durch ein bloßes Hinblicken geschehen kann, sondern nur durch die Erzeugung von Varianten, um so Inhalt und Umfang von Begriffen und die notwendige Verbindung von Begriffen (und den Begriffsgegenständen) anschaulich zu erfassen.

Und was das Verhältnis von Anschauung, Intuition und Urteil betrifft, so kann dies wiederum nur durch Intuition und Deskription in der Reflexion aufgeklärt werden. Schon jedes einzelne prädikative Wahrnehmungsurteil beruht auf Wahrnehmung und präprädikativer Erfahrung als einem bestimmten intuitiven Vollzug am Gegebenen. So setzt das Wahrnehmungsurteil ‚Diese Rose ist rot‘ voraus, dass in der Wahrnehmung eine räumliche Gestalt von bestimmter Form und Farbe gegeben ist und dass diese Gestalt eine bestimmte sinnliche Struktur aufweist: nämlich dass sie eine Ganzheit von Stücken und Momenten ist. Und dieses Verhältnis muss zunächst wahrnehmend erfasst werden, um dann auch zum Ausdruck gebracht werden zu können. Und zwar indem zunächst vorprädikativ die Ganzheit erfasst wird und *dann* der geistige Blick ein Moment der Ganzheit *als* dessen Teilmoment erfasst: „Das unbestimmte Thema S [etwa eine sinnliche Gestalt; Anm. T.S.] wird in der Entfaltung zum *Substrat* der hervorgetretenen Eigenschaften, und sie selbst konstituieren sich in ihr als seine *Bestimmungen*.“ (EU, 126) Erst durch dieses präprädikative Explizieren des Wahrnehmungsgegenstandes wird es möglich ein prädikatives Urteil zu fällen. Die Koplula ‚ist‘ des prädikativen Urteils drückt dabei die ontologische Struktur von Ganzheit und Teilmoment des Gegenstandes aus, die die vorprädikative Explikation als intuitiver Vollzug offenbart hat. Mittels der Koplula und der Prädikate geben wir wieder, wie der Gegenstand partiell ist: rot, blättrig, einen Stil habend etc. Ich muss jedenfalls, bevor ich präzisieren kann, gesehen haben, dass das Dies-da eine Rose ist und dass sie rot ist – und dieses vorprädikative Sehen ist bereits eine geistige (apperzipierend-explizierende) Leistung und kein bloßes Perzipieren. Und es ist die konstitutive Voraussetzung für das Fällen eines Wahrnehmungsurteils.

Der Hauptunterschied zwischen (einfachen) Wahrnehmungsurteilen, empirisch-induktiven und rein-induktiven (also eidetischen) Urteilen besteht darin, dass in den letzten beiden Fällen noch zusätzliche Leistungen erbracht werden

müssen. Im Falle des empirisch-induktiven Urteils müssen mehrere gleichartige Erfahrungen zusammengenommen werden, bei den eidetischen Urteilen muss zusätzlich die eidetische Variation dazwischentreten.

## Zusammenfassung und Ausblick

Ich hoffe hier beispielhaft gezeigt zu haben, dass die Intuition als Akt der Einsicht die Leistung einer grundlegenden Vernunftfunktion darstellt und daher vom Ruch des Irrationalen befreit werden sollte. Der Nous (bzw. die intuitive Vernunft) ist dasjenige geistige Grundvermögen, welches Begriffsbildung und Prädikation ursprünglich ermöglicht. Wenn dies aber richtig ist, dann muss die intuitive Vernunft als eine notwendige Möglichkeitsbedingung der diskursiven Vernunft (und somit schon des Spracherwerbs) verstanden werden. Die obigen Ausführungen legen aber m. E. sogar eine radikalere These nahe: dass es nämlich nicht zwei grundlegende Arten von Vernunft gibt (intuitive und diskursive Vernunft), sondern nur eine Vernunft als eingeborene Disposition, und zwar die intuitive Vernunft, und dass somit die diskursive Vernunft letztlich nichts anderes als eine besondere Anwendungsform der intuitiven Vernunft selbst darstellt. Es ist letztlich die intuitive Vernunft, die durch Sprach- und Wissenserwerb auch zu diskursiven Vollzügen fähig wird.

## Bibliographie

- Aristoteles (1995): *Metaphysik* (übers. v. H. Bonitz). Hamburg: Meiner. (= Met.)
- Cappelen, Herman (2012): *Philosophy Without Intuitions*. Oxford: Oxford University Press.
- Chudnoff, Elijah (2013): *Intuition*. Oxford: Oxford University Press.
- Descartes, René (1993): *Regulae ad directionem ingenii* (übers. v. Lüder Gäbe). Hamburg: Meiner. (= R)
- Kant, Immanuel (2001): *Kritik der reinen Vernunft*. Hamburg: Meiner. (= KrV)
- Kant, Immanuel (2001): *Prolegomena zu einer jeden künftigen Metaphysik*. Hamburg: Meiner. (= Prol.)
- Husserl, Edmund (1999): *Erfahrung und Urteil*. Hamburg: Meiner. (= EU)
- Streubel, Thorsten (2016): *Kritik der philosophischen Vernunft. Die Frage nach dem Menschen und die Methode der Philosophie*. Wiesbaden: Springer VS.

Uwe Meixner

## Truth against Reason, and Reason against Truth

**Abstract:** The paper explores the general relationship between reason (here: epistemic rationality) and truth. It centers on the question of whether there can be conflicts between reason and truth, and on the forms such conflicts may take. Are there philosophically interesting examples of such conflicts?

Rationality is a normative concept, and the language of obligation is as applicable to it as it is to morality. Consider, in particular, epistemic rationality. The language of obligation applied to epistemic rationality enables us to distinguish the following four pairs of prominent *constituents of epistemic rationality*:

1. being rationally *obligated* (O) to believe  $p$ ;<sup>1</sup> being rationally obligated *not* to believe  $p$ ; or in other words:  $O(Bp)$ ;  $O(\neg Bp)$ ;
2. being rationally obligated to believe non- $p$ ; being rationally obligated *not* to believe non- $p$ ; or in other words:  $O(B\neg p)$ ,  $O(\neg B\neg p)$ ;
3. being rationally *permitted* (P) to believe  $p$ ; being rationally permitted *not* to believe  $p$ ; or in other words:  $P(Bp)$ ;  $P(\neg Bp)$ ;
4. being rationally permitted to believe non- $p$ ; being rationally permitted *not* to believe non- $p$ ; or in other words:  $P(B\neg p)$ ;  $P(\neg B\neg p)$ .

There are logical equivalence relations between some of these eight constituents. On the basis of the general schematic principles  $P(A) \equiv \neg O(\neg A)$  and  $P(\neg A) \equiv \neg O(A)$ , we have:

$$P(Bp) \equiv \neg O(\neg Bp) \qquad P(\neg Bp) \equiv \neg O(Bp)$$

---

<sup>1</sup> Note that “ $p$ ” is here a variable for propositions (and not a schematic letter that represents a sentence). As a consequence, the expression “that  $p$ ” is nonsensical (in contrast to the expression “that ( $p$  is true)”: “that”, in the relevant function, must be followed by a sentence, a closed sentence or an open one, not by a singular term. An expression that has the form “that A” is – for any true or false (English) sentence that is substituted for “A” – a name of a proposition, and can be substituted for “ $p$ ”. In what follows, “O(...)” will be treated as a sentence-operator (forming a sentence from a sentence), “B...”, however, as a predicate (forming a sentence from a singular term). For the sake of brevity, the symbol of negation “ $\neg$ ” will be used in a double function: as a sentence-operator ( $\neg A$ : not-A) and as the negation-functor for propositions ( $\neg p$ : non- $p$ ). Both uses occur side by side in  $O(\neg B\neg p)$ , for example.

$$P(B \rightarrow p) \equiv \neg O(\neg B \rightarrow p) \quad P(\neg B \rightarrow p) \equiv \neg O(B \rightarrow p).$$

These equivalence relations allow us to reduce the four *permission*-constituents to the negations of the four *obligation*-constituents. Furthermore, on the basis of the general schematic principle P1:  $O(A) \supset \neg O(\neg A)$ , and its contrapositive P1<sup>c</sup>:  $O(\neg A) \supset \neg O(A)$  (which is logically equivalent to P1), we have:

$$\begin{array}{ll} \text{P1.1} & O(Bp) \supset \neg O(\neg Bp) \quad \text{P1.1}^c \quad O(\neg Bp) \supset \neg O(Bp) \\ \text{P1.2} & O(B \rightarrow p) \supset \neg O(\neg B \rightarrow p) \quad \text{P1.2}^c \quad O(\neg B \rightarrow p) \supset \neg O(B \rightarrow p). \end{array}$$

In other words:

$$\begin{array}{ll} O(Bp) & \text{is contrary to} \quad O(\neg Bp) \\ \text{entails} & \text{entails} \\ \neg O(\neg Bp) & \text{is subcontrary to} \quad \neg O(Bp) \end{array}$$

$$\begin{array}{ll} O(B \rightarrow p) & \text{is contrary to} \quad O(\neg B \rightarrow p) \\ \text{entails} & \text{entails} \\ \neg O(\neg B \rightarrow p) & \text{is subcontrary to} \quad \neg O(B \rightarrow p) \end{array}$$

There are, finally, logical implication relations between some of the eight constituents which, in contrast to the previously considered logical relations, connect “Bp” with “B→p”, and which are, in contrast to the previously considered logical relations, *specific* to epistemic rationality, that is, to the type of rational obligation and permission that concerns propositional belief:<sup>2</sup>

$$\begin{array}{ll} \text{P2} & O(Bp) \supset O(\neg B \rightarrow p) \quad \text{P2}^c \quad \neg O(\neg B \rightarrow p) \supset \neg O(Bp) \\ \text{P3} & O(B \rightarrow p) \supset O(\neg Bp) \quad \text{P3}^c \quad \neg O(\neg Bp) \supset \neg O(B \rightarrow p) \end{array}$$

The eight constituents of epistemic rationality

$$\begin{array}{lll} O(Bp) & \neg O(Bp) & [:= P(\neg Bp)] \\ O(\neg Bp) & \neg O(\neg Bp) & [:= P(Bp)] \\ O(B \rightarrow p) & \neg O(B \rightarrow p) & [:= P(\neg B \rightarrow p)] \end{array}$$

---

<sup>2</sup> It would perhaps have been more appropriate to speak of *doxastic* rationality. On the other hand, the designation “epistemic” emphasizes very appropriately that the rationality in question has something to do with truth.

$$O(\neg B\neg p) \quad \neg O(\neg B\neg p) \quad [:=P(B\neg p)]$$

can be combined in 16 not obviously self-contradictory ways to form *states of epistemic rationality* with respect to the proposition concerned (that is, with respect to the proposition  $p$ ). The *possible* states of epistemic rationality are those that are marked “*consistent*” in the following listing:

O(Bp)	O(¬Bp)	O(B¬p)	O(¬B¬p)	
Y(es)	Y	Y	Y	inconsistent
Y	Y	Y	N(o)	inconsistent
Y	Y	N	Y	inconsistent
Y	Y	N	N	inconsistent
Y	N	Y	Y	inconsistent
Y	N	Y	N	inconsistent <sup>3</sup>
Y	N	N	Y	<i>consistent</i>
Y	N	N	N	inconsistent
N	Y	Y	Y	inconsistent
N	Y	Y	N	<i>consistent</i>
N	Y	N	Y	<i>consistent</i>
N	Y	N	N	<i>consistent</i>
N	N	Y	Y	inconsistent
N	N	Y	N	inconsistent
N	N	N	Y	<i>consistent</i>
N	N	N	N	<i>consistent</i>

In other words,

$$R1: O(Bp) \wedge \neg O(\neg Bp) \wedge \neg O(B\neg p) \wedge O(\neg B\neg p) \text{ i.e., } O(Bp)$$

$$R2: \neg O(Bp) \wedge O(\neg Bp) \wedge O(B\neg p) \wedge \neg O(\neg B\neg p) \text{ i.e., } O(B\neg p)$$

$$R3: \neg O(Bp) \wedge O(\neg Bp) \wedge \neg O(B\neg p) \wedge O(\neg B\neg p) \text{ i.e., } O(\neg Bp) \wedge O(\neg B\neg p)$$

$$R4: \neg O(Bp) \wedge O(\neg Bp) \wedge \neg O(B\neg p) \wedge \neg O(\neg B\neg p) \text{ i.e., } O(\neg Bp) \wedge \neg O(B\neg p) \wedge \neg O(\neg B\neg p)$$

$$R5: \neg O(Bp) \wedge \neg O(\neg Bp) \wedge \neg O(B\neg p) \wedge O(\neg B\neg p) \text{ i.e., } \neg O(Bp) \wedge \neg O(\neg Bp) \wedge O(\neg B\neg p)$$

$$R6: \neg O(Bp) \wedge \neg O(\neg Bp) \wedge \neg O(B\neg p) \wedge \neg O(\neg B\neg p) \text{ [not simplifiable]}$$

are the *six possible states of epistemic rationality* (with respect to the proposition  $p$ , and with respect to a certain – implicit – subject of belief [believer] and a certain – implicit – moment of time).

---

<sup>3</sup> To see the inconsistency in this line, consider that  $O(Bp) \supset \neg O(B\neg p)$  is a logical consequence of P2 and P1.2<sup>c</sup>.



On the other hand, there are *four possible states of belief* (with respect to the proposition  $p$ , and with the respect to the same – the aforementioned – subject of belief and moment of time):

F1:  $Bp \wedge B\neg p$

F2:  $Bp \wedge \neg B\neg p$

F3:  $\neg Bp \wedge B\neg p$

F4:  $\neg Bp \wedge \neg B\neg p$

The possible states of epistemic rationality, as listed above, are consecutively numbered: R1 – R6, and the possible states of belief, as listed above, are also consecutively numbered: F1 – F4. In what follows I will refer to those numberings.

A possible state of belief  $F$  is *rational with respect to* a possible state of epistemic rationality  $R$  if, and only if,  $F$  fulfils all the rational obligations that are intrinsic to  $R$ .

Thus, F1 – that is,  $Bp \wedge B\neg p$  – is not rational with respect to R1, R3, R5, since F1 does not fulfil the obligation  $O(\neg B\neg p)$ , which is intrinsic to R1, R3, R5. And F1 is also not rational with respect to R2, R3, R4, since F1 does not fulfil the obligation  $O(\neg Bp)$ , which is intrinsic to R2, R3, R4. It may seem that F1 is rational with respect to R6. Not so; for  $O(\neg Bp \vee \neg B\neg p)$  is an obligation that is intrinsic to every state of epistemic rationality, since it is purely a matter of the logic of rational obligation;<sup>4</sup> F1 does not fulfil that obligation.

Now, F2 – that is,  $Bp \wedge \neg B\neg p$  – is not rational with respect to R2, R3, R4, since F2 does not fulfil the obligation  $O(\neg Bp)$ , which is intrinsic to R2, R3, R4. But F2 is rational with respect to R1, R5, R6.

In turn, F3 – that is,  $\neg Bp \wedge B\neg p$  – is not rational with respect to R1, R3, R5, since F3 does not fulfil the obligation  $O(\neg B\neg p)$ , which is intrinsic to R1, R3, R5. But F3 is rational with respect to R2, R4, R6.

Finally, F4 – that is,  $\neg Bp \wedge \neg B\neg p$  – is not rational with respect to R1, R2, since F4 does not fulfil the obligation  $O(Bp)$ , which is intrinsic to R1, and not the obligation  $O(B\neg p)$ , which is intrinsic to R2. But F4 is rational with respect to R3, R4, R5, R6.

These results can be perspicuously summed up in a diagram:

---

<sup>4</sup> Note that  $O(\neg Bp \vee \neg B\neg p)$  is logically equivalent to  $O(Bp \supset \neg B\neg p)$ .  $O(Bp) \supset O(\neg B\neg p)$  – that is, P2 – can be taken to follow logically from  $O(Bp \supset \neg B\neg p)$ .

	F1	F2	F3	F4
R1	NR	R	NR	NR
R2	NR	NR	R	NR
R3	NR	NR	NR	R
R4	NR	NR	R	R
R5	NR	R	NR	R
R6	NR	R	R	R

The six possible states of epistemic rationality R1 – R6 and the four possible states of belief F1 – F4 can be combined to form 24 possible *epistemic-rationality-and-fact situations*:  $R1 \wedge F1$ ,  $R1 \wedge F2$ , ...,  $R6 \wedge F4$ . Each of these epistemic-rationality-and-fact situations either has the character “rational” or the character “irrational”. Which one of these two characters it has can be read off the above diagram: What, for example, is the rationality-character of  $R3 \wedge F3$ ? Go to the *row* R3 and the *column* F3, and look where they intersect: There is an “NR” there, and therefore  $R3 \wedge F3$  has the character “irrational”. In the same way it can be determined that  $R4 \wedge F4$  (say) has the character “rational”. It is easily seen from the diagram that ten of the epistemic-rationality-and-fact situations are *rational*, and fourteen *irrational*. Observe also that some possible states of epistemic rationality allow more rational freedom than others (count the “R”s in their rows), and that some possible states of belief are logically more likely to be rational than others (count the “R”s in their columns).

Since R1 – R6 logically exclude each other, and also F1 – F4 logically exclude each other (with respect to the same subject of belief and moment of time, and the same proposition  $p$ ), only one of the 24 possible epistemic-rationality-and-fact situations can be *actual* – with respect to the same subject of belief and moment of time, and the same proposition  $p$ . Which one of the 24 possible epistemic-rationality-and-fact situations will be *the actual one* depends on the proposition  $p$ , on the subject of belief concerned and the relevant moment of time, and, of course, on the pertinent normative facts of epistemic rationality and non-normative facts of belief. Note that only if it is specified which of the 24 possible epistemic-rationality-and-fact situations is *the actual one*, can one speak *simpliciter* of the epistemic rationality or irrationality of a given subject of belief, at a certain time, with respect to a given proposition.

Philosophical positions arise. Consider *radical skepticism*. For radical skeptics, *epoché* is the rationally obligatory epistemic attitude with respect to any proposition. In other words, radical skeptics hold that, for *any* proposition  $p$  and *any* (intelligent)

human subject of belief<sup>5</sup> and *any* time, it is rationally obligatory not to believe  $p$ , and also rationally obligatory not to believe  $\neg p$ . In other words again, radical skeptics hold that R3 – i.e.,  $O(\neg Bp) \wedge O(\neg B\neg p)$  – is the obtaining (or actual) state of epistemic rationality for any proposition  $p$ , no matter what is the time and who (among us human beings) is the subject of belief.

If R3 is the obtaining state of rationality for all propositions, then most of us cannot escape being irrational with respect to some propositions; for most of us cannot avoid believing some propositions, positive and negative ones. I, for example, cannot avoid believing that I exist, and that 2 is not identical with 1. (You may have your own favorites.) Indeed, the rationality of the radical skeptics themselves is undermined by their very position *if it is true*; for they themselves, qua radical skeptics, certainly believe a certain proposition (namely, the proposition that, for all propositions  $p$  and all times, one is rationally obligated not to believe  $p$  and also rationally obligated not to believe  $\neg p$ ).<sup>6</sup> This means that radical skeptics, *if they are right* (i.e., if radical skepticism is true), are bound to be epistemically irrational with respect to the very proposition they qua radical skeptics believe (for the truth of that proposition requires that, rationally, they ought not to believe that proposition – but they do). It is important to note that there is no inconsistency here. If a radical skeptic believes that, for all propositions  $p$  and all times, one is rationally obligated not to believe  $p$  and also rationally obligated not to believe  $\neg p$ , and if it is true that, for all propositions  $p$  and all times, one is rationally obligated not to believe  $p$  and also rationally obligated not to believe  $\neg p$  – then there is no inconsistency in this; then the skeptic is simply right. But he is also epistemically irrational.

The case of the radical skeptic who is right in his skepticism is a particularly striking example of how truth and reason may collide: Someone believes what is, in fact, true; but the logical consequence of his believing what is true is that he is being epistemically irrational (in the sense that he is not fulfilling obligations of epistemic rationality which are incurred by the very fact that what he believes is true). Here is another example of a possible conflict between reason and truth, an example which is much more commonplace than the previous one. Suppose it is rationally obligatory, for any (human) subject of belief, to believe that God does not exist. If so, then R2 is the obtaining state of epistemic rationality with respect to the proposition that God does not exist, and there is only one way for a

---

<sup>5</sup> In what follows it is assumed, for the sake of brevity, that all subjects of belief – in particular, all human beings – are always intelligent.

<sup>6</sup> If with every proposition  $p$  also  $\neg p$  is a proposition (as seems right), the radical skeptics' position can also be expressed by "for all propositions  $p$  and all times, one is obligated not to believe  $p$ ".

subject of belief to be rational about that proposition: this is F3, that is, not to believe that God exists, and to believe that God does not exist. Many people, usually philosophers, have no difficulty at all to comply with the requirements of reason here. But suppose now that God *does* exist – in spite of the (supposed) fact that it is rationally obligatory to believe that God does not exist. This would mean that reason requires one to believe something that is not true. Can true reason counsel, indeed decree, against truth? Can it be rationally obligatory to believe something that is not true?

One is tempted to say “no” and to postulate the following principles that connect the rational obligation to believe a proposition with the truth of that proposition:

P4  $O(Bp) \supset p$  is true [i. e.,  $\neg p$  is not true]

P5  $O(B\neg p) \supset \neg p$  is true [i. e.,  $p$  is not true]. [pair 1]<sup>7</sup>

These postulates may even be strengthened:

P6  $O(\neg B\neg p) \supset p$  is true [i. e.,  $\neg p$  is not true]

P7  $O(\neg Bp) \supset \neg p$  is true [i. e.,  $p$  is not true]. [pair 2]<sup>8</sup>

The first pair of postulates follows from the second pair in view of the uncontroversial principles

P2  $O(Bp) \supset O(\neg B\neg p)$

P3  $O(B\neg p) \supset O(\neg Bp)$

that have already been introduced and made use of. And one can argue for the second pair in the following manner: Can it be rationally obligatory not to believe something even though it is true? Can true reason counsel against truth? It seems not. Thus, if reason says “Thou shalt not believe that  $p$ ” and is right,  $p$  has to be *not* true, and non- $p$ , therefore, true. And if reason says “Thou shalt not believe that non- $p$ ” and is right, non- $p$  has to be *not* true, and  $p$ , therefore, true.

---

<sup>7</sup> Pair 1 obviously entails  $\neg O(Bp) \vee \neg O(B\neg p)$ , or in other words:  $O(Bp) \supset \neg O(B\neg p)$  – which is an unproblematic principle that is also a consequence of P2 in combination with P1.2<sup>c</sup>:  $O(Bp) \supset O(\neg B\neg p)$  and  $O(\neg B\neg p) \supset \neg O(B\neg p)$ .

<sup>8</sup> Pair 2 obviously entails  $\neg O(\neg Bp) \vee \neg O(\neg B\neg p)$ , or in other words:  $O(\neg Bp) \supset \neg O(\neg B\neg p)$  – which, if true, would for *every* proposition render skepticism (or agnosticism) impossible as an obligation of epistemic rationality; for according to  $O(\neg Bp) \supset \neg O(\neg B\neg p)$ , R3 cannot be an obtaining state of epistemic rationality for any proposition.

Principles P4 – P7, if accepted, have interesting consequences. The moral obligation to do something, or not to do something, does not entail that it is done in fact, or not done in fact. Correspondingly, the rational obligation to believe, or not to believe, a certain proposition does not entail that it is believed in fact, or not believed in fact. But, if P4 is accepted, the rational obligation to believe a certain proposition entails that it is true, and if P7 is accepted the rational obligation not to believe a certain proposition entails that it is not true. There is no parallel of this on the side of morality; it is peculiar to epistemic rationality.

What are other consequences of accepting P4 – P7? For one thing, it becomes quite impossible to justly accuse people of being irrational because they believe a certain proposition. For example, the atheist or agnostic cannot with justice tell the believer that she is being irrational by believing that God exists. For the believer can legitimately defend herself by answering that there is no rational obligation for her (or anybody) not to believe that God exists, and therefore no rational obligation to believe that God does not exist, either. Why? Because, says the believer, it is just not true that God does not exist.<sup>9</sup>

The dilemma is this: On the one hand, the rational obligation to believe  $p$  is to be a *good epistemic indicator* of the truth of  $p$ . For that seems to be the whole point of the rational obligation to believe a certain proposition. On the other hand, the truth of  $p$  is to be a *conditio sine qua non* of the rational obligation to believe  $p$ . For it would seem that true reason cannot decree against truth. But if the rational obligation to believe  $p$  is to be a good epistemic indicator of the truth of  $p$ , then the truth of  $p$  cannot also be a *conditio sine qua non* of the rational obligation to believe  $p$ . For if the truth of  $p$  were a *conditio sine qua non* of the rational obligation to believe  $p$ , then the rational obligation to believe  $p$  would be drawn into question already by a sincere claim that  $p$  is not true – and indeed already by a mere uncertainty whether  $p$  is true. Clearly, under such circumstances the rational obligation to believe  $p$  cannot be a good epistemic indicator of the truth of  $p$ .

Consider a similar case, but a case where there is no dilemma. The truth of  $p$  is a *conditio sine qua non* of *knowing* that  $p$  is true; one cannot know that  $p$  is true without the truth of  $p$ . But this makes it quite impossible for knowledge to be a good epistemic indicator of truth. If one is uncertain whether  $p$  is true, one cannot reasonably say to oneself: “Person X certainly knows that  $p$  is true; this indicates that  $p$  is true.” For the uncertainty whether  $p$  is true is ipso facto an uncertainty whether anybody *knows* that  $p$  is true – since the truth of  $p$  is a *conditio sine qua*

---

<sup>9</sup> She is using the contrapositives of P7 and P3 (i.e., P3<sup>c</sup>) to arrive at her conclusions:  $\neg(\neg p \text{ is true}) \supset \neg O(\neg Bp)$  and  $\neg O(\neg Bp) \supset \neg O(B\neg p)$ .

*non* for knowing that *p* is true. But of course one *can* reasonably say “It is true. I know it to be true” or “It is true. N.N. said that he knows it to be true”. The first is merely an emphatic way of saying “I believe it to be true”, and the second is merely a compendious way of saying “N.N. believes it to be true, and I share his belief because N.N. is in this matter an authority for me”.

In resolving the above-presented dilemma for rational epistemic obligation – which dilemma is a consequence of conflicting conceptual aims – we should, after all, *not* follow the example set by knowledge. The similarity between *rational obligation to believe* and *knowledge* is far from being strong. Saying that it is rationally obligatory (for subject X) to believe *p* is just another way of saying that there cannot be any reasonable doubt (for X) about *p*'s being true.<sup>10</sup> But how can it be the case that there cannot be any reasonable doubt about *p*'s being true if this impossibility of reasonable doubt can only be the case if *p* is true? If the truth of *p* were a *conditio sine qua non* of the rational obligation to believe *p*, then, *except* for the *epistemically special* propositions (the uncontroversially true propositions in logic and mathematics),<sup>11</sup> it would never be rationally obligatory to believe *p*; for one must always allow (*except* in the case of an epistemically special proposition) that the believed or surmised non-truth of *p* can reasonably be held against the alleged rational obligation to believe *p*.

But if the truth of *p* is not in general a *conditio sine qua non* of the rational obligation to believe *p*, then, indeed, in a considerable number of cases the claim that there is a rational obligation to believe a certain epistemically non-special proposition may be true. We may even come to the point that 99.9% of such claims are true: that in 99.9% of the cases in question there is indeed, as is claimed, a rational obligation to believe a certain epistemically non-special proposition – for example, just for the sake of the argument, that God does not exist, that everything is physical, that there is no incompatibilist freedom of the will, that life has no ultimate meaning, that there are no objective moral obligations. But, note, the claim that it is, e. g., rationally obligatory to believe that God

---

**10** Saying that it is *not* rationally obligatory (for subject X) to believe *p* is just another way of saying that there *can be* reasonable doubt (for X) about *p*'s being true. Saying that it is rationally obligatory (for X) *not* to believe *p* is just another way of saying that there *is* reasonable doubt (for X) about *p*'s being true.

**11** The proposition *that I exist* is another epistemically special proposition – a very special one. It is rationally obligatory to believe that I exist – but it is so only *for me*: there is no rational obligation for anyone else (who is human) to believe that I exist. I cannot reasonably doubt (the truth of) the proposition that I exist, but anyone else (who is human) *can* reasonably doubt that proposition. In contrast, it is rationally obligatory for anyone (who is human) to believe the proposition that  $1+1=2$ ; no one (who is human) can reasonably doubt that  $1+1=2$ .

does not exist is now in a peculiar sense *fallible*. By this I do not merely mean that the obligation-claim itself may be false (I think that it is false, but I have just now proposed, for the sake of the argument, that it is true); I mean that the claim, though true, may be nothing more than an imperative of reason *against truth*.

Consider, finally, a telling parable. By accidental contact with human beings, the idea that there are human beings has touched a population of intelligent *termites*; call the termites in that population “the alpha-termites”. Yet, in the course of their history (in which further contact with human beings is for a very, very long period non-occurrent), the alpha-termites ultimately come to the following claim: There cannot be any reasonable doubt that there aren’t any human beings. In other words, it is rationally obligatory, for every alpha-termite, to believe that human beings do not exist. And this obligation-claim is even true: It is, indeed, rationally obligatory, for every alpha-termite, to believe that human beings do not exist. How could it be *not* rationally obligatory for an alpha-termite to believe this, given that the truth of a proposition, as was finally determined in this paper, is *not a conditio sine qua non* of the rational obligation to believe it?<sup>12</sup> But, as we human beings *know*, the imperative of termite-reason in question is an imperative *against truth*. The alpha-termites, by the way, never learned the truth; they were destroyed by human beings before they had any chance to learn it – which goes to show that there are certainly worse things than being irrational: *being wrong* is one of them.

## Appendix

(I) *Principles in full explicitness*: With time-index and subject-index added,  $P2 - O(Bp) \supset O(\neg B\neg p)$  – (for example) turns into:  $O_{s,t}(B_{s,t}p) \supset O_{s,t}(\neg B_{s,t}\neg p)$ . If also the quantification involved is made fully explicit, P2 turns into  $\forall s\forall t\forall p(O_{s,t}(B_{s,t}p) \supset O_{s,t}(\neg B_{s,t}\neg p))$ .

---

<sup>12</sup> Still, readers may wonder *why* there cannot be any reasonable doubt – for the alpha-termites – that there aren’t any human beings. The savants of the alpha-termites would tell potential doubters among the alpha-termites (and the savants would tell the truth): (1) There is not a shred of dependable evidence for the assumption that there are human beings; (2) every phenomenon in the (alpha-termite) world can be perfectly explained without assuming that there are human beings – and *entia non sunt multiplicanda praeter necessitatem*; (3) it is even “incoherent” to assume that there are human beings, for human beings “just don’t fit in”. This may leave one with the impression that it is merely rationally obligatory for every alpha-termite *not to believe* that *there are* human beings – and not that it is rationally obligatory for it *to believe* that *there aren’t any* human beings. But the difference which looks big in logic is very small in practice – *given* the truth of (1), (2), and (3).

(II) *If radical skepticism is true, then the radical skeptic is irrational:*

---

(1)	$\forall s \forall t \forall p (O_{s,t}(\neg B_{s,t}p) \wedge O_{s,t}(\neg B_{s,t}\neg p))$	[thesis of radical skepticism]
(2)	$q^* := \text{that } \forall s \forall t \forall p (O_{s,t}(\neg B_{s,t}p) \wedge O_{s,t}(\neg B_{s,t}\neg p))$	[definition]
(3)	$B_{(s^*,t^*)}q^* \wedge \neg B_{(s^*,t^*)}\neg q^*$	[s* is a radical skeptic at t*]
(4)	$O_{s^*,t^*}(\neg B_{s^*,t^*}q^*) \wedge O_{s^*,t^*}(\neg B_{s^*,t^*}\neg q^*)$	[instantiation of (1)]
(5)	s* is irrational at t* with respect to q*	[from (3) and (4)]

---

Here the actual state of epistemic rationality is of the type R3 and the actual state of belief is of the type F2; therefore the actual situation-of-rationality-and-fact is of the type  $R3 \wedge F2$  – which is irrational, since F2 is irrational with respect to R3. Replacing (3) by (3') –  $B_{(s^*,t^*)}q^* \wedge B_{(s^*,t^*)}\neg q^*$  – does not help. The only way to escape irrationality in the face of (1) is replacing (3) by (3''):  $\neg B_{(s^*,t^*)}q^* \wedge \neg B_{(s^*,t^*)}\neg q^*$ . But if this is the actual state of belief of s\* at t\*, then s\* can hardly be called “a radical skeptic”; otherwise, the famous “man on the street”, who never thinks about epistemological matters, would turn out to be a radical skeptic.

(III) *Prima facie it may seem that “p is true  $\supset$  O(Bp)” (the converse of P4) is a valid principle.* However, if p is true and it is nevertheless impossible to believe p (for example, because p just cannot be grasped by the subject of belief) then there is no obligation to believe p. Likewise, if p is true and p is totally irrelevant for the subject of belief, then there is no obligation to believe p. Thus, “p is true  $\supset$  O(Bp)” is certainly not a valid principle; but other valid principles are implicit in the considerations that show its invalidity, for example:  $O(Bp) \supset \diamond(Bp)$ .







# Philosophy of Mind



Jan Slaby

# ***Don't beep me, bro!* – A Worry About Introspection<sup>1</sup>**

**Abstract:** I argue against the feasibility of certain employments of introspection as a method in psychological experience research. The chief mistake that dooms these attempts is a deep-seated misconstrual of human experience, namely, the omission or under-appreciation of the role of *agency* in experience. Adopting a line of thought that is prominent in Wittgenstein's *Philosophical Investigations* and that was later elaborated by Richard Moran, I argue that self-constitutive agency is the linchpin of a viable understanding of the human mind. I then illustrate how missing out on this essential aspect of our mindedness hampers one prominent recent attempt to revive introspection as a method in psychology – Russ Hurlburt's experience sampling method. In closing, I hint at some broader considerations concerning the dangers involved in scientific objectifications of the mind.

## **1 Introduction: Return to the fly bottle?**

Wittgenstein famously said that his aim in philosophy was “to show the fly the way out of the fly bottle” (PU, § 309). His philosophical aspiration was to surpass a problematic mentalism that lets each of us be engaged in our own minds. Wittgenstein undertook it to liberate us from a pervasive and pernicious framework of thought that would lead to skepticism about other minds and to skepticism about knowledge in general – a framework centered around the idea that all individual thinkers ever have mental access to is their own inner representations. It is easy to slip back into a version of this framework, as the many exemplary episodes and ways of speaking that Wittgenstein discusses in his later writings abundantly testify. Modern thought in general is prone to fall into the Cartesian trap, which consists in inappropriate ontologisations of an important insight, namely that individual perceptions, thoughts or judgments might be mistaken and thus stand in constant need of rational scrutiny. To conclude from the constant possibility of error that there must be *ontic mediators* between thinking

---

<sup>1</sup> As the present paper grew out of a conference presentation, I have kept some of the informal spoken style of in the hope that this might rekindle in the reader some of the performative force of what was a partly polemical, discussion-provoking oral presentation.

subjects and objective reality is the master mistake in this area (cf. McDowell 1995).<sup>2</sup>

The worry that I will point to in the following is that some lines of empirical work in psychology and the cognitive sciences may unwittingly contribute to leading Wittgenstein's fly back into the bottle. In its eager attempt to capture conscious experience in its "full qualitative richness", this kind of work is in acute danger of *misconstruing the mind* – so that it ultimately amounts to a giant leap backwards for philosophy as well as for science, fields that have by and large been liberated, quite laboriously, from the worst of their mentalistic obsessions. Accordingly, I will revisit some of the reasons for why the mental cannot be empirically captured by introspective self-observation in the way required for scientific research, regardless of how "qualitative" and how subtle and refined one's experimental methods and set-ups might be. There is something flawed, or at any rate highly problematic in the very idea of first personal *description* of mental life.

I will discuss my worry by bringing it into relatively sharp relief, i.e. I will deliberately state both the views I critique and my own alternative in relatively stark, rigorous and explicit terms. My hope is that the clarity of the contrast will outweigh the losses in fine detail.<sup>3</sup>

## 2 An innocent-seeming assumption

The well-meaning philosopher reflecting about experimental work in cognitive neuroscience might come up with the following advice: A scientific investigation of mind-enabling bodily mechanisms had better find a role for careful, first-personal characterizations of mental states, for only then can researchers be sure to be targeting the right kinds of explananda for their empirical investigations. The aim is to capture mental states as they unfold so as to get a better grip on their material foundations. It is a highly non-trivial task to individuate, and authenti-

---

<sup>2</sup> The following is among Wittgenstein's most pointed statements in this domain of his thought: "Wenn wir sagen, *meinen*, dass es sich so und so verhält, so halten wir mit dem, was wir meinen, nicht irgendwo vor der Tatsache: sondern wir meinen, dass *das und das – so und so – ist*." (PU § 95)

<sup>3</sup> A more detailed and somewhat more balanced elaboration of the problem space I address in the following is the excellent chapter on methodology in Shaun Gallagher's and Dan Zahavi's landmark book *The Phenomenological Mind* (2008). Among its many merits is a detailed demonstration that the founding fathers of phenomenology, notably Husserl and Merleau-Ponty, were quite innocent of – and themselves arduous critics of – the conceptual and methodological flaws that are discussed in the present paper.

cally characterize, human mental states or mental processes. Failing to do so would doom the sciences of the mind from the outset, because of descriptive inadequacy – researchers would simply look for the wrong things. Accordingly, let experimental subjects *introspect* and self-report on their introspection. For this to succeed, it is crucial to give subjects the time and resources needed to elaborate, in enough detail and precision, on their experiences. Make use of the rich repertoire of phenomenology – and maybe meditation and mindfulness techniques and other established methods – to refine subjects' first-personal access to their experience, and – if possible – train your subjects in these modes of access, endow them with the right kinds of concepts and techniques of self-observation so as to reap from them maximally accurate descriptions. With these enriched descriptions in hand, investigators stand much better chances to be on target in their search for the mind-enabling machinery.<sup>4</sup>

One striking example is the so-called experience sampling method developed by the psychologist Russell Hurlburt (see, e.g., Hurlburt & Heavey 2006; Hurlburt & Schwitzgebel 2007; Hurlburt 2009). Experimental subjects are provided with an electronic device – a beeper – that will emit a beeping sound at various random times during a day, upon which subjects are required to immediately halt what they are doing and instead write down as accurately as possible what was going on “on their minds” in the very moment that they heard the beeper go off. As it is rather difficult to introspect accurately in this manner, Hurlburt and his collaborators work hard at training their subjects so as to make them less prone to mistakes during self-observation. All too easily, according to Hurlburt, subjects “contaminate” their introspective reports with retrospective interpretation, background knowledge or temporal misattribution. With the right kind of training, so the thinking goes, will subjects be increasingly able to report accurately on what Hurlburt tellingly calls “pristine experience” (Hurlburt 2009).

Doesn't this sound reasonable and innocent enough, and doesn't it make good sense from an empirical standpoint? What can possibly be wrong with approaches like this? I think a lot goes wrong here, and quite fundamentally at

---

<sup>4</sup> A number of different proposals can be found in the literature. Much of it was kick-started by Francisco Varela and his naturalized phenomenology movement (see, e.g., Varela et al. 1991; Depraz et al. 2003; Varela & Shear 2009) A small step removed from the Varela roots but still carrying echoes of it are Jack & Roepstorff (2003), Petitmengin (2006) and Bitbol & Petitmengin (2013). No discernibly background in the naturalized phenomenology tradition has Russell Hurlburt, whose proposal is a particularly apt target for my present critique (see Hurlburt & Heavey 2006; Hurlburt 2009).

that. Thus the title of the present paper, which expresses my central message to Hurlburt, his team, and to all like-minded researchers: *Don't beep me 'bro!*<sup>5</sup>

### 3 Two worries about introspective research

The problem with this introspective research strategy is that its starting point is still fundamentally Cartesian, regardless of the often-voiced materialist conviction that mind is physically realized in bodily mechanisms. One key assumption – sometimes tacit, sometimes, as in Hurlburt, explicit – is that the mind is an individual experiential space, accessible first-personally, that in some as of yet unknown way maps onto processes in the brain and organism of the individual. On this assumption, each of us is the star witness to the goings-on in a private, inner experiential space. Accordingly, something like introspection has to be the method of choice, it is the only real avenue to access what is of interest here.

Two related things are worrisome here. First, this understanding of the mind leaves in place a great portion of the original Cartesian framework, viz. the inward retreat of experience. The first-personal descriptive strategy assumes a mental space, separated from the rest of the world, to which each of us has exclusive, or in any case privileged access. This is fundamentally unsatisfying for several reasons. This framework notoriously invites skepticism, and it seems to fly in the face of much that seems evident about our predicament as situated worldly beings.

Second, and more importantly for present purposes, this understanding of the mind construes the relation each of us has to their own mind in terms of a *passive* stance of *bearing witness* to mental processes and contents. It assumes that it is an authentic predicament to be an observer of oneself, even an “inner” observer, beholding mental goings-on on an inner stage or arena of experience. This foregrounding of a *spectatorial relationship* as a model for subjective experience and self-knowledge is fundamentally misleading – a pernicious picture that still keeps some of us captive. Crucially, such a picture of the mind is at

---

<sup>5</sup> This alludes to a slogan first yelled in desperation by a student in Florida moments before police uses a *taser* to restrain him: *Don't tase me bro!* The slogan quickly went viral and created some buzz, ending up on t-shirts and elsewhere (see, e.g. <http://www.wired.com/2007/09/dont-tase-me-br/>; visited on June 2, 2015). Obviously, part of my point is that the assumptions about the human mind build into Hurlburt's experience sampling research design are not altogether outside the ambit of a mind-control-fantasy with a ring of thought police. Views of the human mind are usually laterally entangled with visions of the political – a fact that many present-day philosophers of mind are quite good at blocking from view.

odds with the all-important way in which our *agency* is implicated in our mental lives – or so I will argue in the following.

## 4 The centrality of agency

The fact that we are agents is fundamental to an appropriate understanding of the human mind.<sup>6</sup> Wittgenstein himself said it succinctly, namely that “*im Anfang war die Tat*” (ÜG, 402). *Agency* in the soon-to-be-specified sense is the one focal concept that we cannot omit or reduce, it is quite simply the deepest and most essential component of our self-understanding as mindful creatures.

The mental is inextricable from agency, in the strong, ontological sense that the mental is itself fundamentally *agentive*. This is not meant in the weak sense that mental contents are *de facto* often acted upon, and not in the even weaker sense that there often is a kind of spontaneous mental *activity* going on in us, but in the strong sense that what we have “on our minds” is in fact in important parts *shaped* – often even literally produced – *in the act*, i.e. in, as and by our own doing. In crucial respects, our relationship to our own mental states – if we allow us to speak in this objectifying manner – is one of agency rather than one of observation. Much of our self-consciousness comes in a package with agency: it is the activity of making up one’s mind and the possibility of acting, and being called upon to act, so as to resolve conflicts, assume a stance, make a decision, resist or give way to an impulse, follow or break off a line of thought, keep oneself composed, attentive, on the ready, and so on. Observation – being a witness of one’s mental life – is secondary at best to this fundamentally active, future-oriented dimension.

I adopt this centrality-of-agency position for the most part from Richard Moran’s immensely illuminating elaboration of it (in his 2001 book *Authority and Estrangement*). Moran’s point is that we gain an adequate understanding of the human mind when we place a certain form of agency at the center of the account. To have a mind is *to be* a mind, and to be a mind is to be a rational agent – an agent undertaking and living up to epistemic commitments, such as beliefs, and likewise undertaking and living up to conative commitments, such as desires. Thus, the agency in question, when it comes to what gets usually called “mental states”, is chiefly the *making up of* one’s mind about what is true (the result will be our beliefs) and likewise *making up of* one’s mind about what is worthy of our pursuit (the result will be our desires). Accordingly, mental

---

<sup>6</sup> In this section, I substantially draw on considerations developed by Richard Moran (2001).



states are not objective items in an inner theater of some sort, but the various *active stances* adopted by a rational agent.

Take the case of being asked what you believe about some non-trivial matter X. Do you respond to this question by looking inside yourself, to see what it is that you believe about X? No, you rather look into the world, toward X, and make up your mind about it – you deliberate, for instance by actively recalling what you know about X, by judging and weighing options – thereby you shape, actively, your attitude towards whatever is at issue. And when you avow your belief, you thereby undertake a *commitment* that includes, among other things, a readiness to defend your view when challenged, and to assent also to what follows from your belief, and so on. It is not a matter of just *finding yourself saddened*, passively, with ready-made mental contents. A self-ascription about a matter X meets what Richard Moran, echoing Gareth Evans, calls the *transparency condition* when it is made by consideration of the facts about X itself. Belief is transparent to the world in this way; it is the ability to let one's epistemic stance be controlled by what is the case (cf. Moran 2001, p. 92–93).

When we act and decide to resolve conflicts, we *make up our minds* on what it is we *should* think, judge, or feel in a given situation, we adopt a stance towards things that we're ready to live up to against obstacles, we *commit* to a future course of action, actively hold on to it or revise it in the light of new experiences, exercise guidance with regard to our intending and acting in line with what we intend. Take the case of an intention to act: in the salient cases, we do not *observe* an intention just welling-up passively within us; rather, we *form* the intention, sometimes through conscious deliberation, but always by way of holding on to a thought, making it relevant for what we are going to do, taking and living up to a stance of intending. Surely, there might be an element of mere impulse or inclination in the formation of an intention, but we have an important say in letting it determine or affecting the course of our action. Likewise, memory: we actively remember, make up our minds about some past event, over and above merely “recalling” past sequences of experience. There is an element of mere recall, and we depend in our remembering on things popping into our mind, but the act of remembering is centrally an active stringing together of a past episode, the activity of relating, the act of assuming a stance to what once was and an active readiness to work it into shape for the present encounter. And certainly thought: *thinking* is in an important sense an activity, a piecing together of thought contents, drawing out relevant connections, taking a firm stand on an issue, and also one of holding back distracting influences, actively focusing on a topic, keeping one's attention at the right spot, and so on.

Even emotion is in no small part a matter of agency: think of being in a difficult social situation – is it so out of the question to ask yourself what you *should feel* in this situation, and then actively resolve the issue by *committing* to a specific emo-

tional response, i.e. making a scene, bursting into anger, displaying visible anxiety, or deflating a moment of awkward tension with a joke, etc.? But even in less confusing or ambivalent situations, emotion – with the exception of primitive reflex emotions such as panic, surprise etc. – usually is dependent upon the agent's *continuing* an initial impulse by way of committing to an emotion's evaluative and agentive trajectory, to keep enacting the emotion, acting in line with its evaluative and practical orientation. Thereby, the situation and one's own stance in it as a participant is actively shaped in the emotion, a certain situational value is crystallized, made and kept manifest (cf. Slaby & Wüschner 2014).

Of course there *are* predominantly passive mental goings-on – sensations such as itches, stabs, tickles and of course pains are paradigmatic –, and also cases such as “being struck by a thought”; but even here: we don't just register these neutrally and leave it at that; rather, we encounter a sensation with an attitude, we adopt a stance towards it. There is a “take-it-or-leave-it”-moment even with regard to initially passive mental occurrences. It makes a crucial difference to the sensation how we let it affect us, or how we take a thought that has “struck us” into our strides, how we go on rolling with it (cf. Helm 2001, ch. 3).

In a word: Agency is inextricable from the constitution of the kind of mentality that we can credit ourselves with. This all too easily gets lost on a self-observation model of mental access. Without us in this way active, called upon to make up our minds, to resolve conflicts, adopt or sustain attitudes, to constantly clear up the “mess inside” (Goldie 2012) as best we can, actively unify diverging currents of experiences etc., there would be no mind as we know it. The highly problematic metaphor of a *stream of consciousness* is half-way adequate only if we acknowledge that this stream is not a subject-less transpiring, but something that comes with a keeper, a guardian – an instance that keeps house, as it were. The agent needs to be in the picture, on pains of missing out on what is distinctive about minds like ours.

That is also why ascriptions of mental contents to a person always come with a limited warrant, as there is never a fixated, definite content on any person's mind as long as the agentive core is in operation. This is because every assumed content that becomes mentally manifest is by default subject to active shaping, to dismissal or transformative endorsement – the mind is no chamber, field or screen on which definite “items” make their appearance. Instead, it is the domain of doing – an ever-changing, active process. As Richard Moran writes with a nod to Sartre: “no attitude or impulse apprehended by reflective consciousness has any right to continued existence apart from one's free endorsement of it” (Moran 2001, p. 140). So whatever mental structures, drives, or traits of character I am confronted with in my self-consciousness, in becoming conscious of them I'm both in the position to distance myself from them – I must

be able to see myself as *distinct* from these various facticities in my consciousness – and I must be able to temporarily *bracket off* their psychic reality, I can disengage from them so that it is now *up to me* to either newly endorse them, or fight them, work against them, or acquiesce in them, or some such. To be sure, this does not mean that I am free to choose whatever character traits or psychic features I would like to possess – the view expressed here is not a crude version of voluntarism. Rather, it means that as long as I am a reflectively self-conscious agent, nothing I become conscious of will drift entirely free of my endorsing, dismissing, engaging or disengaging with it. To be conscious in the way persons are conscious is to be a self-responsible agent with regard to whatever one becomes conscious of – and this is true even in those situations where what one becomes conscious of is an overwhelming facticity that one has little choice but to “put up with” in one way or other. Nothing I become conscious of is a foregone conclusion insofar as I am to adopt an attitude toward it (cf. Moran 2001, p. 140).

## 5 An aside on the limits of empathy

The centrality of agency for our mental lives likewise tends to get lost or downplayed on models of interpersonal interaction that build on simulation or empathetic perspective shifting (i. e. those “higher” forms of empathy that are sometimes called cognitive empathy as they require the complex maneuver of projecting *oneself* into the shoes of another; see Coplan 2011 for clarification): what can exactly *not* be simulated in empathetic interaction is the agency at the heart of the other person’s perspective on the world – the way the person commits herself, takes a stand, bestows evaluative weight on the momentary options she has, seizes opportunities, resolves conflicts among discordant thoughts or feelings, and so on. The problem with empathetic perspective-shifting is that the one who is shifting perspective can only ever bring her *own* agency to bear (Goldie 2011; cf. Slaby 2014). Trying to simulate another’s mental states is a kind of foreign imposition, a move close to patronizing the other because one inevitable will take what is in fact *one’s own* agency (or would-be agency) for the agency in which the other person’s mental states are constitutively anchored. As agents, we are in a crucial sense “irreplaceable”; agency is in each case essentially someone’s, there is an ineliminable moment of authentic ownership, or *Je-Meinigkeit*, to use Heidegger’s term (Heidegger 1927, p. 41).<sup>7</sup>

---

<sup>7</sup> In Slaby (2014) I have developed these skeptical considerations with regard to the prospects

Thus we see that this line of thought reaches back to the existentialist tradition, especially to Sartre and Heidegger. Being an agent means quite literally *being it yourself*, as nobody can take over your agency and decide and live in your stead. Following Joseph K. Schear, we can put a still slightly different gloss on this crucial thought and frame it in the following way: being a human subject means “*being able to fail at being a subject*” (Schear 2009, p. 105; emphasis added). Subjectivity in the full human sense is an existential task, something that has to be actively undertaken and actively kept up as long as one continues to exist as a person. There is a constant doing, a constant holding-it-all-together and composing oneself, keeping oneself on track, at the ready, in the here-and-now – on pain of, quite literally, “loosing oneself”.<sup>8</sup>

## 6 Situating the agent

The agentive perspective on the mind points towards a way to resolve the first worry I expressed above – that of the inward retreat of experience into a peculiar “mental space”. As we have already seen, the predicament of the agent is utterly misdescribed if we construe it as a case of dealing with mental contents – let alone with contents that are fixated like material items, or contents individually purified, separated from one another and stripped of their entanglements with the world. Much rather, the correct picture is closer to this one: The agent has to be seen as *situated in the midst of worldly circumstances* – circumstances that present him or her with opportunities, obstacles, dangers, solicitations, affordances – invitations or demands to act. Accordingly, when the task is to understand someone, it is advisable to ask people not about what’s going on “in them”, but rather about their situation, their lives, about what goes on *with*

---

of empathetic perspective shifting in more detail, drawing on Moran and Goldie. There, I also make clear that these critical considerations will not doom every understanding of empathy. For instance, phenomenological interaction theory – based on the direct sharing of a situation, often situations of joint agency – might offer a much more promising route to an understanding of empathy that does not involve the vexed operation of mental perspective-shifting as in “imagining-being-the-other”. On interaction theory, see Gallagher (2008) and Krueger (2011).

**8** That to be a subject means to be able to actively fail at being one leads back in a straight line to Aristotle’s extremely rich concept of *hexis*. Some of the ways in which this concept can help to better capture the involvement of agency in our mental and personal lives are tentatively explored, with a focus on the emotions, in Slaby & Wüschner (2014). It may well be that this specific way of holding-oneself, of actively keeping one’s perspective on the world in play and in shape, is exactly that which gets catastrophically eroded in cases of severe depression (see Slaby et al. 2013 and Ractliffe 2015).

and *around* them – how they are effectively *situated* in the world, wrapped-up in their place, as it were. Thereby, we bring into view a broader practical context – the concrete possibility space of the agent instead of a subjective realm of experience. Only a focus on the situation in which a person finds herself brings the agency aspect of the mental properly into view – the existential fact that we are called upon to act in or respond to a particular situation that we are currently embedded in. So in the end, we do not inquire about a person’s mind, but about a person’s *life* – a stance much more adequate to the kinds of active and situated beings that we are.

This brings the view that I have sketched here in line with the theory of enaction in the philosophy of mind and in cognitive science (see, e.g., Varela et al. 1991; Noë 2004; Thompson 2007). Enactivism assumes a constitutive interrelatedness of organism and environment, which presumably grounds a “deep continuity” of mind with processes of life (Thompson 2007). Crucially, the term “enaction” refers to the specific process mode in which an organism, as part of its self-organizing and self-preserving metabolic activity, meaningfully relates to its surroundings – a relatedness whose working principle is not representation, but *activity* in the form of viable sense-making: effective conduct aimed at sustaining and promoting the organism’s identity as a living being. Transposed into the realm of adult human mentation, this means that what it is to be a mindful being is to be understood in terms of activity *in* the world, not in terms of alleged inner representational states that are *about* the world. The verb “to enact” is employed in a double sense, taking two different classes of objects – what is enacted is both what is on the object side of a relating activity (world, a domain, a realm of significance), and also the subjective mode of conduct, the mode of relating to the object: consciousness, perception, mind etc. are enactive, i.e. they unfold as meaningful *activities* in the world. Enaction thus establishes a fundamental mutuality: the agent and her agentive conduct on one side, and a specifically enacted practical domain, an agent-specific environment on the other side. For present purposes, obviously, the biological dimension of this, the dynamic organism-environment mutuality, is of lesser importance. What is central is the understanding of the mental as fundamentally agentive. Mental activity, according to enactivism, is viable conduct in the world; being a minded creature is nothing other than to continuously self-sustain and realize oneself in the act.<sup>9</sup>

---

<sup>9</sup> Much of this has been already developed in sophisticated ways by Merleau-Ponty, whom Evan Thompson in particular credits as a crucial precursor to enactivism (see especially Merleau-Ponty 1942).

## 7 The wrong angle on the mind

Of course there *is* something going on “in us”, we all also just *have* certain mental states, we all find ourselves with mental goings-on that we’re not bringing about directly by our own conscious agency. So I am certainly not claiming that it is *completely absurd* to focus inwardly, to introspect, or even to refine our conceptual capacities in this regard, so as to learn to discern patterns and structures that would not be noticeable without a specifically focused or even trained self-observation. And of course it might be possible to do empirical research on this. My point is this: What we thereby reveal provides at best a one-sided, limited angle on the human mind. Just a fraction of what is key about being a human agent is brought in view in this way, by focusing on isolated mental episodes. One may put the point even stronger: If we take the passively generated, the naturally occurring, the “merely observable” as our paradigm mentality, we in fact come close to framing or construing a different mind from the one that comes in view when we take agency as the kernel of the mental. This observational mind is a mind without agent, without owner, without responsibility – a realm of agent-less unfoldings, of mere happenings. It is not much more than a series of isolated snapshots; the all important processual core, the heart of the matter is quite simply left out.

## 8 Is the experimentalists’ mind made-up?

There is a particular danger here. Some experimental uses of introspection – and many experimental settings in general – actively encourage a suspension of agency. There are many experimental designs that urge subjects to refrain from acting, from actively interfering in the course of their experience, in order to just behold what “happens in them”. Many experimental settings limit or delimit agency. One might say that the experimenter *actively* puts his subjects into a *passive* position. This is quite pronounced in Hurlburt’s work, where the experimentalist is at pains to stop subjects from “contaminating” their carefully prompted self-reports from all sorts of actively added ingredients, such as self-interpretation, reflection, comparisons with previous mental episodes, and so on. Or think of many fMRI experiments in cognitive or affective neuroscience: It is hard to imagine a more restrictive, more agency-inhibiting situation than lying still in a brain scanner. Nonetheless, work based on fMRI experiments has given rise to dominant and ever-expanding research paradigms, for instance in social and affective neuroscience – fields that have been close

to the forefront of a massive scientific reframing of personhood. If the mind-as-agency thesis as defended here is even half-way correct, these approaches effect the opposite of what they aim at. Instead of accessing the mental “as it is” (surely a problematic idea in and of itself!), they introduce a highly artificial element of passivity. They suspend what is most crucial about minds like ours – *our self-constitutive agency* – and help build up what is indeed a partially objectified, passive mental space.

So the risk is, in effect, that we contribute to *producing*, in experiments making use of introspection, the very passive, self-objectifying subjects that are assumed on those problematic mentalistic theories. That this is even possible is an effect of our being *self-interpreting animals* (Taylor 1985). We can *make* specifications of ourselves *true* by taking them at face value, by, as it were, implementing them in our modes of conduct and ways of life (that is what it means to be able to *commit* to stances, to ways of being etc.). It is a paradoxical, or, at any rate, highly ironical situation: The fact that the passivist mentalistic picture is misguided – the fact that our mind is *exactly not* a ready-made inner sphere of passively generated contents, but rather a matter of self-constitutive agency – comes strikingly to the fore in a process of shaping our mental lives whose outcome may exactly be the initially wrong picture: a passive, self-obsessed, objectified mind – a mind that stops considering itself as the (partial) author and sustainer of itself, stops considering itself as the instance called-upon to actively ensure the unity, coherence and direction of a dynamic perspective on the world. This is possible, because a deep truth about our minds is that we can *indeed* shape ourselves, actively, in all sorts of existential directions, so potentially even in the direction of a highly artificial passivity. In case the institutional, discursive and technological environment that supports this framing and shaping is robust and powerful enough, as might be the case with a dominant strand of the experimental human sciences, the self-constitutive agent might be considerably impeded, potentially even up to the point of vanishing entirely.<sup>10</sup>

## 9 The passivized mind – A product of our times?

This points to a fundamental worry familiar from recent critiques of the “psychologistic complex”: i. e. that the PSY disciplines – and their various customers and pro-

---

<sup>10</sup> Fortunately, this assessment might be a little premature, as indications point in the direction of the global neuro-hype of the past two decades coming to a halt (see Slaby 2015 and Slaby & Gallagher 2015 for more on this).

moters – *create* their subject matter much more than they reveal something truly pre-existing (see, e.g. Brinkmann 2008; de Vos 2012; Slaby 2010). This pertains also to the “self” of the “self report” in many experimental approaches in psychology and cognitive neuroscience. How can we prevent dealing with an experiential structure that we ourselves have framed and ushered into being in our experimental set-ups? Let me stress one thing about this: what is thereby created might be absolutely real. It is not that we just deal with a bunch of conceptual confusions that might have little effect, as austere adherents of ordinary language philosophy sometimes tend to assume. Instead, we have to read constructivist proposals in an eminently realist way. When the material and discursive surround is robust enough, then what is conceptually constructed can be as real as anything, and so even conceptual confusions can have eminently real consequences.<sup>11</sup>

This has wider ramifications.<sup>12</sup> Could it be that today’s culture at large is increasingly calling forth a passive, consumerist, self-objectifying self – so that what we are doing in the neuroscience or psychology labs is congenial to what is going on in globalized, capitalist societies, in companies, universities, administrations? Is the self-conscious, reflective, responsible agent on the way to become a relic from a bygone era – a relic from a time where there still *were* significant cultural spaces for individual agency? Whereas in our time – a time where bosses are flourishing – autonomy and rational self-determination are on the way to becoming idle academic fictions: less and less a real option in the socio-technological landscapes that make up today’s lifeworlds – and thus likewise less and less real as a fundamental characteristic of ourselves. After all, it might well be that the self of the self-report in introspective psychology and cognitive neuroscience is “neuronal man”, the ideal template of 21<sup>st</sup> century subjectivity. Every era has the science it deserves.

## Bibliography

Bennett, Maxwell R./Hacker, Peter M. S. (2003): *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.

---

**11** In part, this goes against the grain of Bennett & Hacker’s (2003) otherwise impressive Wittgensteinian critique of the neurosciences, in which the authors sometimes seem to imply that nothing much in the way of “real consequences” can come out of the usual assortment of conceptual confusions that neuroscientists are prone to get tangled up in. I do not think this is so, but I cannot discuss this point any further here.

**12** The following is the barest of sketches; I say more on this line of thought in Slaby (2015).



- Bitbol, Michel/Petitmengin, Claire (2013): "A defense of introspection from within". In: *Constructivist Foundations* 8(3), pp. 269–279.
- Brinkmann, Svend (2008): "Changing psychologies in the transition from industrial society to consumer society". In: *History of the Human Sciences* 21(2), pp. 85–110.
- Coplan, Amy (2011): "Understanding Empathy". In: Amy Coplan/Peter Goldie (Eds.): *Empathy: Philosophical and Psychological Perspectives*. Oxford: Oxford University Press, pp. 3–18.
- Depraz, Natalie/Varela, Francisco, & Vermesch, Pierre (2003): *On Becoming Aware*. Amsterdam: John Benjamin's.
- De Vos, Jan (2012): *Psychologisation in Times of Globalisation*. New York: Routledge.
- Gallagher, Shaun (2008): "Direct Perception in the Intersubjective Context". In: *Consciousness and Cognition* 17, pp. 535–543.
- Gallagher, Shaun, & Zahavi, Dan (2008): *The Phenomenological Mind: An Introduction to Philosophy of Mind and Cognitive Science*. London & New York: Routledge.
- Goldie, Peter (2012): *The Mess Inside: Narrative, Emotion, and the Mind*. Oxford: Oxford University Press.
- Goldie, Peter (2011): "Anti-Empathy". In: Amy Coplan/Peter Goldie (Eds.): *Empathy: Philosophical and Psychological Perspectives*. Oxford: Oxford University Press, pp. 302–317.
- Heidegger, Martin (1927 [1986]): *Sein und Zeit*. Tübingen: Max Niemeyer.
- Helm, Bennett W. (2001): *Emotional reason. Deliberation, motivation, and the nature of value*. Cambridge: Cambridge University Press.
- Hurlburt, Russell/Heavey, Christopher L. (2006): *Exploring inner experience*. Amsterdam: John Benjamin's.
- Hurlburt, Russell/Schwitzgebel, Eric (2007): *Describing inner experience: Proponent meets sceptic*. Cambridge, MA: The MIT Press.
- Hurlburt, Russell (2009): "Iteratively Apprehending Pristine Experience". In: *Journal of Consciousness Studies* 16, No. 10–12, pp. 156–188.
- Jack, Anthony/Roepstorff, Andreas (Eds.) (2003): *Trusting the Subject I*. Special double issue of *Journal of Consciousness Studies* 10/9.
- Krueger, Joel (2011): "Extended cognition and the space of social interaction". In: *Consciousness and Cognition* 20 (3), pp. 643–657.
- McDowell, John (1995): "Knowledge and the internal". In: *Philosophy and Phenomenological Research* 55(4), pp. 877–893.
- Merleau-Ponty, Maurice (1942): *The structure of behavior*. Pittsburgh: Dusquene University Press.
- Moran, Richard (2001): *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton UP.
- Noë, Alva (2004): *Action in Perception*. Cambridge, MA: MIT Press.
- Petitmengin, Claire (2006): "Describing one's subjective experience in the second person: An interview method for the science of consciousness". In: *Phenomenology and the Cognitive Sciences* 5(3–4), pp. 229–269.
- Ratcliffe, Matthew J. (2015): *Experiences of Depression: A Study in Phenomenology*. Oxford: Oxford University Press.
- Schear, Jonathan K. (2009): "Experience and Self-Consciousness". In: *Philosophical Studies* 144, pp. 95–105.

- Schwitzgebel, Eric (2008): "The unreliability of naïve introspection". In: *Philosophical Review* 117(2), pp. 245–273.
- Slaby, Jan (2015): "Critical Neuroscience meets Medical Humanities". In: *Medical Humanities* 41(1).
- Slaby, Jan (2014): "Empathy's Blind Spot". In: *Medicine, Health Care and Philosophy* 17(2), pp. 249–258.
- Slaby, Jan/Paskaleva, A./Stephan, Achim (2013): "Enactive Emotion and Impaired Agency in Depression". In: *Journal of Consciousness Studies* 20 (7–8), pp. 33–55.
- Slaby, Jan (2010): "Steps towards a critical neuroscience". In: *Phenomenology and the Cognitive Sciences* 9, pp. 397–416.
- Slaby, Jan/Gallagher, Shaun (2015): "Critical Neuroscience and Socially Extended Minds". In: *Theory, Culture and Society* 32(1), pp. 33–59.
- Slaby, Jan/Wüschner, P. (2014): "Emotion and Agency". In: S. Roeser/ C. Todd (Eds): *Emotion and Value*. Oxford: Oxford University Press, pp. 212–228.
- Taylor, Charles (1985): "Self-Interpreting Animals". In: *Human Agency and Language. Philosophical Papers Vol.1*. Cambridge: Cambridge University Press, pp. 45–67.
- Thompson, Evan (2007): *Mind in Life. Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard UP.
- Varela, Francisco/Shear, Jonathan (1999): "First-person Methodologies: What, Why, How?". In: *Journal of Consciousness Studies* 6(2–3), pp. 1–14.
- Varela, Francisco/Thompson, Evan/Rosch, Eleanor (Eds.) (1991). *The Embodied Mind*, Cambridge, MA: MIT Press.
- Wittgenstein, Ludwig (1951 [1984]). *Tractatus-Logico-Philosophicus, Tagebücher 1914–1916, Philosophische Untersuchungen*. Werkausgabe Bd. 1. Frankfurt am Main: Suhrkamp [abbreviated as PU].
- Wittgenstein, Ludwig (1970): *Über Gewißheit*. Frankfurt am Main: Suhrkamp. [= ÜG].



James N. McGuirk

# Metaphysical and Phenomenological Perspectives on Habituality and the Naturalization of the Mind

**Abstract:** It is a commonplace amongst neuroscientists and philosophers to consider habit to be a ‘naturalization’ of the mind, in which ways of thinking and ways of acting sink below the level of conscious reflection and become thoughtless, blind, uncritical and unthinking. This paper seeks to challenge that view, not by rejecting the idea of habit as naturalization of the mind, but by challenging the received view of what such naturalization entails. Drawing on the work of Felix Ravaisson, and more specifically phenomenologists Maurice Merleau-Ponty, Paul Ricoeur and Gaston Bachelard, the paper argues that close attention to the role of embodiment and temporal unfolding in habitual acting demand a more nuanced understanding of habit as well as of the relationship between mind and nature generally.

In a recent article on the neuroscience of habits in *Scientific American*, the following standard test for establishing the presence of habits is described:<sup>1</sup>

A group of rats is placed in an experimental box, and trained to press a lever, which releases food as a reward. When the rats are returned to their cages, the experimenters ‘devalue’ the reward in the box either by arranging it such that the rats can eat to oversatiation, or by injecting them with a drug that induces nausea when they eat the food. When the rats are then returned to the experimental box, the presence of a habit is identified on the basis of whether the rat continues to press the lever and eat the food or not. If they do so, in spite of the fact that the reward is now unpleasant, they are said to have formed a habit. If, on the other hand, they do not, they are said to be ‘mindful’ of the alteration in the situation and are said not to be acting habitually.

Researchers then studied the neurological processes in play in the rats acting habitually in order to gain insight into the neuroscience of habits generally. While there is much of interest that can be said about the details of this research, it is not something we can go into in any detail here.

What is striking, however, is that this experiment or versions thereof, remains the standard way in which habits are assessed in neuroscientific research

---

<sup>1</sup> Research carried out at the Graybiel laboratory at MIT (Graybiel & Smith 2014)

on the topic. It is interesting because it reveals several key assumptions about the nature of habits. These include:

1. that habits involve actions which are carried out “thoughtlessly”
2. that this thoughtlessness is so comprehensive that even negative outcomes will not alter habitual behaviors (or do so only very slowly)
3. that habits are actions triggered by external stimuli and involve unchanging responses to these stimuli.
4. that the behavior of rats in an experimental box is for all intents and purposes adequate to drawing conclusions about the habitual behavior of human beings.

These assumptions about habit are not, however, explicit hypotheses to be tested. They are rather taken for granted as truisms precisely because they are so deeply embedded in ordinary discourse on the nature of habit, something which in its turn is at least as indebted to the filtering down of philosophical theories of habit as it is to ordinary experience. Spinoza considered habits the expression of intellectual laziness, while Kant identifies them with moral torpor. David Hume sees in habituality the persistent enactment or thinking of what has been learned either through rote, custom and/or the contingent organization of experience. For all three, habituality is expressive of action or thinking which is uncritical, unthinking, and disengaged.

In the 20<sup>th</sup> century, Gilbert Ryle is acutely wary of habits in spite of his commitment to forms of knowing which are not explicitly articulate. Ryle famously argues against reducing intelligence to propositional knowledge and claims that intelligence is predominantly characterized by dispositions to act intelligently. Intelligence is not first and foremost identified with explicit “knowing that” but with embedded “knowing how”. But while rejecting the idea that the proposition has a monopoly on the intelligent, he simultaneously rejects the idea that habit has anything to do with intelligence at all. Dispositions are tendencies to act intelligently, says Ryle, while habits are the result of drill and manifest no intelligence on the part of the agent. Take the example of the multiplication tables. The tables themselves are the product of intelligent organization, but their recitation tells us as little about the presence of intelligent agency as does the parrot’s recitation of a line from *Hamlet*. The kind of drill involved in habit “dispenses with intelligence”, rather than developing it (Ryle 2000, p. 42) because what it teaches is automatic responses to identical external stimuli.

Like Aristotle, Ryle calls habits “second natures” but like Gendler, he considers these second natures to be “associative, automatic and arational” states, which are triggered by external stimuli and which we share with non-human an-

imals (Gendler 2008, p. 641).<sup>2</sup> It is clear, that he considers this naturalization of the mind as its dissolution. To act from habit, he says, “is to act automatically and without mind to what one is doing” (Ryle 2000, p. 42). The difference between the habit and the disposition is that while the disposition involves intelligent action which may seem thoughtless, the habit involves thoughtless action which seems intelligent. This is because even to the extent that the habitual action indicates intelligence (such as the recitation of the multiplication tables), the performance itself is devoid of intelligent engagement.

When it comes to habit, in other words, Ryle is very much a behaviorist. And while Graybiel and Smith do not explicitly identify themselves as behaviorists, it is clear that their assumptions clearly are. And perhaps they have a point. These behaviorist assumptions do, after all, contain a good deal of intuitive reasonableness.

We often discuss habitual actions as thoughtless responses and use them to explain the inexplicable. I turned right at the juncture even though I knew that my destination was to the left. Why? Because, I usually take the right turn towards Dublin and wasn't thinking about what I was doing. I was acting from habit, we say. The same road or stimulus, the same action or response. I may even repeat this mistake time after time because the action is like a reflex that takes over unless I consciously arrest it. The habit tends to repeat even after it has served its initial purpose, i.e. the benefit of freeing mental space for us to attend to other matters. This accounts for the assumption of Ryle and the neurologists, that habit is *most* in play when it results in counterproductive performance. The rats, we remember, were said to act habitually not because they persisted in pressing the lever and eating the food but because they did so *in spite of* the deleterious effects of doing so. They pressed the lever, to begin with, because of an expected outcome. Now, they press it simply because it is there. In other words, habit, by its nature, involves a non-intelligent engagement with one's environment which is why it is predominantly in the light of negative results that the presence of the habit can be definitively confirmed. This does not mean, of course, that habitual action always yields negative results but only that the formation of habits intrinsically weakens the capacity of the subject to respond mindfully and with discernment to their environment.

In other words, the view of habit as arational does capture an important aspect of our ordinary experience of the habitual. However, I want to suggest that there is more to habit than the examples given would suggest. The examples, such as that of the rat in the box, or recitation of the multiplication tables are

---

<sup>2</sup> This description actually refers first and foremost to what Gendler dubs *aliefs*, although she claims later in the paper that what goes for the *alief* goes also for the habit (Gendler 2008, p. 663).

not useless but they invite hasty conclusions about habits as simply thoughtless and, by extension, of the idea that the naturalization of the mind entailed by habit should be understood as the loss of mind to the logic of cause and effect.

I want to consider certain suggestions that challenge the second of these assumptions such that they subsequently transform our view of the first of them.

I will first outline an alternative understanding of what is at stake in the naturalization of the mind than the one we have seen so far. Here I will refer to Felix Ravaisson's short book *On habit* (Ravaisson 2009). I will then show how insights from this work are developed in Merleau-Ponty's account of habit such that not only an alternative understanding of naturalization is offered but also an alternative account of its emergence. To conclude, I will look at suggestions in treatments of habit by Paul Ricoeur (2007) and Gaston Bachelard (2013) which connect the notion of habit not only with a genuine way of knowing but also with invention and creativity.

## 1 The alternate view

While Ryle and other purveyors of the negative view of habit often refer to habits as 'second natures' after Aristotle, they do not think of habit in the spirit of Aristotle, for whom it was an essential aspect not only of the moral formation of humans but of human intellection generally. The second nature, for Ryle, is understood as the loss of the rational to nature. When employed by Felix Ravaisson, on the other hand, the notion of the 'second nature' is decidedly more positive. For him, habits are not only important moral categories, but are crucial to the formation of our knowledge of the world. Habit formation entails a naturalization of the mind, but he insists that this must be understood otherwise than through the logic of cause and effect. It is placed in the context of the relationship between freedom and nature or spontaneity and compulsion, but in a way that challenges the tendency to think these dualistically. Habit is explicable in terms of neither pure freedom nor pure nature. It is instead the intersection of the two. What he says is that mind sinks down into nature in the formation of habit, which is to say that nature – both in general and in terms of the being of the human agent as natural entity – is transformed by the presence of mind. As he puts it,

In descending gradually from the clearest regions of consciousness, habit carries with it light from those regions into the depths and dark night of nature. Habit is an acquired nature, a *second nature* that has its ultimate ground in primitive nature, but which alone explains the latter to the understanding (Ravaisson 2008, p. 59).

Ravaissou acknowledges here a certain opacity in the operation of habit. But the reason for this is not that intelligence has given way to blind impulse but that the once explicit initiative has become embedded in the action that is no longer initiated by an act of explicit thought. Habits may become seemingly effortless, he says, but they do not thereby “withdraw from the...intelligent activity from which they were born” (Ravaissou 2008, p. 57). In other words, while habit does involve the receding of mind from acting, we should understand this as the transformation of the natural by the rational rather than as the loss of the rational to the natural. In the formation of habit, mind is deposited in nature as opposed to being erased by nature. Ravaissou even goes so far as to claim that the habit over time becomes indistinguishable from instinct but in this way is to be understood as revealing the “intimate essence and necessary connection” (Ravaissou 2008, p. 67) between necessity in nature and the freedom of the will. So, while habit may involve the recession of deliberate mental acts, this recession is not to be understood as the flight of reason but as the extension of the mind beyond the deliberative mental act.

According to the Rylean taxonomy, habits comprise non-intelligent degenerations of mind which stand opposed to the genuine intelligence entailed in propositional “knowing that” and dispositional “knowing how”. For action or thought to become *natured* is to become habituated and to become habituated is to degenerate. For Ravaissou, by contrast, it is the natural trajectory of mind to sink down into nature which is to say that habit is not the other of intelligent thought and action but its destiny. At this point, the disagreement is metaphysical in nature inasmuch as Ravaissou freely acknowledges the tendency of attention to recede in habitual acting but denies that mind natured in this way is mind lost. He reverses Ryle, we might say, in claiming that rather than mind becoming eroded by nature, we should understand nature as giving way to mind in this process of naturalization.

## 2 The phenomenological tradition

But the metaphysical is not the only possible means of challenge, as is clear from Merleau-Ponty’s phenomenological treatment of habit. Like Ravaissou, Merleau-Ponty is interested in exploring the relationship between mind and nature (Merleau-Ponty 1983, p. 2) and, like Ravaissou, finds in habit, an exemplary instance of the interface between the two. But the former’s phenomenological emphasis on embodiment means that we find here an entirely different kind of contribution to the discussion of habit formation.



Merleau-Ponty's focus on embodiment opens for a new way of understanding the way in which habits are *acquired* and not just the way they are *enacted*. That meaning is first and foremost a bodily phenomenon, that is, means that habit is not simply about depositing the will into nature, such that the latter is transformed over time, but is equally a case of the presence of nature in the will. He insists that the habituation of skills and indeed the habituation of world engagement is bodily from the very beginning. The body is not primarily an object for an "I think" (Merleau-Ponty 2012, p. 155), it is not an instrument that is commanded by acts of intellection but is the original locus of our encounter with the world as the site of meaning.

This in turn calls for a new understanding of the role of reflection in habit. Merleau-Ponty is undoubtedly sympathetic with the neo-Aristotelian account of habit found in Ravaisson but rejects his prioritization of active reflection for the exercise of intelligence and therefore for the assessment of habits *as* intelligent. For Ravaisson, the habit is first and foremost the result of an act of the will which over time recedes from conscious awareness, such that the habit is *enacted* as a kind of post-reflective performance and draws its intelligence from the original reflective act.

Merleau-Ponty, by contrast, thinks of habits as primarily formed in the absence of such explicit reflection. In other words, habits are neither ontologically nor chronologically dependent upon acts of reflection but are rather enacted intelligences that primarily emerge pre-reflectively. Habits are not only, then, once explicit or intentional actions, which over time become routine and recede from explicit consciousness. So Merleau-Ponty might be understood as agreeing with Ryle against Ravaisson that habits can be formed independently of acts of reflection while claiming with Ravaisson against Ryle that this is no obstacle to considering them intelligent.

He does not deny that certain habits are the result of initially explicit acts of will but he insists that the majority of our habits are the result of the body's power of "dilating our being in the world" (Merleau-Ponty 2012, p. 145). That is to say, that the body is able to incorporate significances into itself in order to transform both itself and its environment. He says at one point that;

the body has understood and the habit has been acquired when the body allows itself to be penetrated by a new signification, when it has assimilated a new meaningful core (Merleau-Ponty 2012, p. 148).

He offers several well-known examples to illustrate this point which include the blind man's incorporation of his cane in sensing the world, the (somewhat quaint) well-dressed lady's adaptation to her feathered hat in navigating exter-

nal space and even the skill of typing which when learned is no longer the result of explicit acts of willing the motion of the fingers (if it ever was this) but “the incorporation of the space of the keyboard into [...] bodily space” (Merleau-Ponty 2012, p. 146). It is in this sense that he speaks of the body as the site of intelligence, which is able to adapt, respond and transform itself through its environment and its environment through it.

There are two crucial points worth mentioning here which constitute the difference between Merleau-Ponty’s view and that of Ravaisson. The first is that he emphasizes the dialectic between mind and nature that comes to view most pre-eminently in the habituality of the body. That is, there is a mutual transformation, of the mind by the world and the world by the mind and not just a one-way sinking of mind into nature.

The second important point has to do with the very notion of the mind *as* embodied. For Merleau-Ponty, as for Ravaisson, the habitual act is one in which will and explicit reflection are peripheral but this is not to say that *consciousness* is peripheral. The habitual action is consciously intelligent *as* embodied, which not only means that the habit is still considered intelligent, as Ravaisson had already claimed, but also that the intelligence is more robustly given to the agent first personally, as the felt sense of the intelligence of the act. In other words, the intelligence of the action is not the trace of the now forgotten act of the will but emerges *in and as* the body’s interface with nature. This means that even the opacity that Ravaisson was willing to acknowledge needs to be re-evaluated. Habitual actions are opaque only with regard to explicit reflection. When I drive, I may no longer think about shifting gears or adapting to the road surface but this is not to say that I am not attentive. Rather the attention which is present is predominantly an embodied attentiveness. I attend to what I do as the body’s attunement, which is also to say that it is as this attunement that consciousness is present in acting habitually. There are hints that Ravaisson anticipates Merleau-Ponty on this point, but it is nevertheless first with the latter that the idea of habit as both intelligent and emergent prior to reflective acts is explicitly suggested.

These analyses of the notion of habit pose a significant challenge to the assumptions with which we began and which continue to dominate both ordinary and specialist discourse about what habits are.

For example, it seems illegitimate to think of habitual action as thoughtless (claim 1). The fact that we do not think explicitly about what we are doing when driving or typing does not mean we are not thinking at all. It does not even mean that the thinking has receded but that, following Merleau-Ponty, the thinking is enacted bodily.

Likewise, the dialectic between body and world that lies at the core of habit formation as described by Ravaisson and especially Merleau-Ponty makes the

behavioristic assumption (claim 3) about habit problematic. Neither would deny that habituated action entails the naturalization of mind but reject the crude equation of naturalization with physicalist mechanisms of cause and effect. The habitual act is not simply a stock response to a stock stimulus but involves a dynamic incorporation of the possibilities to act into the schema of the body. And while the example of the rat in the experimental box is not entirely without value (claim 4), even in assessing habituality in humans, it is far too prejudicial as a test case to provide any significant insight into habit as such, given that the concept of habit covers an enormous range of behaviors, from the seemingly involuntary tic to complex behaviors such as the habit of critically engaging with the prime minister's speeches. This complexity is too often belied in the choice of paradigm example, which does more to uncover the crypto-Cartesian prejudices built into the way we speak and think about habits than it does about the matter itself.

For Merleau-Ponty, as for Ravaisson, the point is to draw attention to the fact that habits are dynamic and enacted in complex and changing environments such that experimental designs which fix environmental conditions once and for all provide little insight. In fact, it is arguable that the unchanging conditions in these experiments *invite* the thoughtlessness that tends to develop, even in rats, as opposed to being merely against which such thoughtless habituation develops.<sup>3</sup>

### 3 The temporality of habit

But while we have dealt with claims 1, 3 and 4, we still have the problem of accounting for claim 2, which had to do with the tendency of habitual action to misfire regarding outcomes. The aforementioned critiques are important but we must avoid the danger of presenting a view of habit that simply replaces the negative view with one that is excessively positive. We noted earlier that automatizations in habitual acting are not only a part of the way we speak about habits, they are also experientially resonant. And while it is perhaps wrong to make the misfiring habits the sole locus of the discussion, the phenomenologists still have the problem of explaining how these misfires are possible.

A clue to solving this problem emerges out of two other discussions of habit in the phenomenological tradition. These are Paul Ricoeur's *Freedom and Nature* and *The Intuition of the Instant* by Gaston Bachelard. What distinguishes these discus-

---

<sup>3</sup> Empirical support for this idea can be found in Bruce Alexander's "rat park" experiments and his work on addiction (Alexander 2010, p. 193–5)

sions from Merleau-Ponty's is that they explicitly connect the notion of habit with that of temporality. This is important because it both allows us to account for misfiring habitual actions while simultaneously advancing our understanding of what is at stake in the dynamism of habitual acting.

A clue to understanding how temporality is relevant is interestingly enough to be found in the paradigm case with which we began. There is a temporal dimension in this experiment, to be sure, in the sense that it is assumed that the habit is formed when the action that was carried out in the past is repeated into the future. However, as we have said, it is reasonable to question the extent to which these experimental conditions are adequate in recreating real life conditions. It is more likely that the inertia or thoughtlessness which comes to characterize the behavior of the rats is thoughtless precisely because they have been conditioned to act in a context in which nothing changes. The experimental design is, in other words, simply a poor one in achieving what it sets out to achieve.

But if this is so, how is it possible that habitualities stagnate in the way that they sometimes do? According to both Ricoeur and Bachelard, it is because the habit becomes disproportionately rooted in the past to the point that the actor becomes closed to the possibilities given in the unfolding present. In other words, the habituated action can only maintain itself in a context, which both unfolds and also disrupts past experience.<sup>4</sup> It must renew itself in time in order to survive.

Ricoeur's discussion of habit in *Freedom and Nature* is heavily influenced by both Merleau-Ponty and by Ravaissou and in combining them, he extends the scope of the phenomenological analyses into a more comprehensive account of habit in human life. On the basis of his hermeneutic sensibilities, Ricoeur anchors the phenomenology of habit in the temporal unfolding of a human life in general. He is concerned not only with bodily habits (Merleau-Ponty), but with intellectual habits (Ravaissou), which may be the result of either once present acts of knowing or may owe their origin to the tradition to which I belong and to which I am given over.

What is interesting from our point of view, though, is the focus in Ricoeur's text not on the past behind the habit but the way in which it opens the future for us. "To acquire a habit", he says "does not mean to repeat and consolidate but to invent, to progress" (Ricoeur 1966, p. 289). Thus, it involves opening the world before me as offering unique possibilities to act and to think.

---

<sup>4</sup> This is similar to the way William James speaks about habituality as plastic enough to be altered while not losing its integrity as its structure changes. (James 2007, 104–5).

The examples given are, as always, telling. While Ryle speaks of the rote recitation of the multiplication tables which indeed involves something close to a thoughtless repetition, and Merleau-Ponty speaks of typing which demands a more nuanced understanding of habit as an incorporation of what is external to the body into the body's schema, Ricoeur uses the examples of the athlete, the orator and the writer (Ricoeur 1966, p. 302).

Such examples underline the need to view habit as predominantly flexible (Ricoeur 1966, p. 284) and inventive. The athlete's movements, like the writer's pen, seek to find new ways of enacting what they have incorporated so that they invent every bit as much as they repeat. The habit reproduces itself in terms of a productive imagination that emerges out of the world behind the text of experience but which is primarily concerned with discovering possibilities for action in the world in front of that text. So habit is held in the tension between repetition and novelty – not something fixed but an unfolding that is only maintained in its flexible and dynamic exercise.

This point is actually anticipated in Merleau-Ponty's famous and oft cited example of the football player's perception of the playing area. According to Merleau-Ponty, the player does not experience the field as an object but as a network of "openings which call for a certain mode of action" (Merleau-Ponty 1983, p. 168)

This example has important implications for the way we think about habits. When the football player engages the field as lines of force, he is specifically engaged in a practical species of thinking that engages the field as a field of possible actions. The game has rules which mean that the lines and spaces have a certain meaning within that context. However, these demarcations do not impel action but invite it and they invite it by opening for a range of *possible* engagements. The football player's habit gives rise to a "probing", as Ricoeur puts it (Ricoeur 1966, p. 290), which co-creates the meaning of the space in the dialectic of transforming and being transformed, such that the agent meets the pitch as a field of possible action which can be engaged imaginatively.

This engagement must, of course, be anchored in the past of the habit as acquired. Neither Ricoeur nor Bachelard, as such, ignores the centrality of repetition for the enactment of habits, but they insist that the habit must be maintained by a repetition that is attuned to the novelty that each situation presents.<sup>5</sup> This repetition is one that "restructures even as it rediscovers itself" (Bachelard 2013, p. 44). We must desire progress if the habit is to preserve its efficacy (Bachelard 2013, p. 45), so that the football player who does not seek to play better today than he did yester-

---

<sup>5</sup> A comparable idea can be found in Bergson's thoughts on the perpetuity of the novel in experience (Bergson 2007, p. 77)

day will play worse tomorrow (Bachelard 2013, p. 38). Ricoeur calls habit “a new structuring in which the meaning of elements changes radically” (Ricoeur 1966, p. 288) while Bachelard calls it the “routine assimilation of novelty” (Bachelard 2013, p. 37). The habit is then transformative of what lies before us and is not only the result of transformations of what lies behind us. This means, in turn, that the habit which does not seek to transcend its previous limits will not simply degenerate into automatism but will dissipate.

Habit is fragile and the actor can easily lose sight of what is before her and give herself over to a repetition, which is *mere* repetition. The habit then slides into automatism. We become buried under habits, and resign our freedom to the weight of custom, of the “only natural”, and so on (Ricoeur 1966, p. 301).

This problem of inertia is a constant threat to habitual acting. But inertia is only possible as a degeneration of habit in the sense of a degeneration in the felt tension between past and unfolding present such that we precisely fail to respond to the situation in which we find ourselves. The threat of ossification is always inscribed in habit but it is not its normal destiny.

For both Ricoeur and Bachelard, the time of the subject is crucial in the analysis of habit in the sense that the unfolding of a life entails the givenness of a present which is both continuous and which subtly breaks with the past. This attention to the temporal is entirely absent from Ryle’s discussion and while it is not incompatible with Merleau-Ponty’s discussion, it was not central there either. Habitual action builds upon the past, of course, but it is enacted in the novelty of the present instant. The instant, says Bachelard, presents a new aesthetic configuration (Bachelard 2013, p. 37; p. 43) which calls for an attuned response that modifies and transforms our capacity for action. But this means that habit cannot be understood as either stock response, or even as a modification of the body or mind that has been achieved. In very suggestive language, Bachelard says that “a habit is a certain order of instants chosen from the basic ensemble of moments in time (that) plays itself out at a specific pitch and with a distinct tone” (Bachelard 2013, p. 43). So habit is not something achieved but the process of achieving such that it dies if it does not continue to grow, to modify itself, and to discover itself and the world anew at each instant.

## 4 Conclusion

The various phenomenological texts, which we have discussed provide important challenges to fundamental assumptions about the notion of habit as thoughtless, blind, uncritical, and unthinking. They do so on the basis of analyses of embodiment and the temporal unfolding of habitual acting which de-

mand a more nuanced understanding of habit as well as of the relationship between mind and nature.

If these accounts are credible, they are important for guiding the way in which habits of mind and body are studied such that we can be alive to the degeneration of habit without misidentifying this degeneration as its essence.

The habit lives, only so long as it continues to progress and discover new possibilities for action. None of this means that habits are returned to the realm of pure self-transparency. In the same way that bodily habits emerge pre-reflectively, the habits of the mind escape me to the extent that there is, as he says, an “it thinks” present in the “I think” (Ricoeur 1966, p. 294). In other words, the life of consciousness is not perfectly transparent to itself. It is, again in this sense, *natured* in the unfolding of its time. However, it remains present to itself as long as it brings its past to bear on the demands of its present situation. That is, it is free *and* naturalized. The inertial problem or the problem of the degeneration in habit occurs precisely when this attention to detail which habit makes possible is absent.

## Bibliography

- Alexander, Bruce K. (2010): *The Globalization of Addiction*. Oxford: Oxford University Press.
- Bachelard, Gaston (2013): *The Intuition of the Instant*. Evanston: Northwestern University Press.
- Bergson, Henri (2010): *The Creative Mind*. New York: Dover.
- James, William (2007): *Principles of Psychology*, vol. 1. New York: Cosimo.
- Gendler, Tamar Szabó (2008): “Belief and Alief”, in *Journal of Philosophy* 105 (10): 634–663
- Graybiel, Anne & Kyle Smith (2014). “Good habits and bad habits”, in *Scientific American*. 310 (6): 38–43.
- Merleau-Ponty, Maurice (2012): *Phenomenology of Perception*. London: Routledge.
- Merleau-Ponty, Maurice (1983): *The Structure of Behavior*. Duquesne: Duquesne University Press.
- Ravaisson, Felix (2009): *On Habit*. London: Continuum.
- Ricoeur, Paul (1966): *Freedom and Nature*. Evanston: Northwestern University Press.
- Ryle, Gilbert (2000): *The Concept of Mind*. Middlesex: Penguin.

Thomas Fuchs

# Embodied Knowledge – Embodied Memory

**Abstract:** The distinction between representational and embodied knowledge (knowing-that versus knowing-how) has gained new significance through the investigation of implicit memory. This kind of memory is formed in the course of the interaction of organism and environment: Recurring patterns of interaction are sedimented in the form of sensorimotor, but also affect-motor schemes. We may speak of an implicit “body memory” that underlies our habits and skills, connecting body and environment through cycles of perception and action. This embodied knowledge is actualized by suitable situations or by overarching volitional acts, without necessarily being made explicit.

The paper analyses the structure of embodied knowledge by taking the example of learning social skills through dyadic interactions in early childhood. It argues that the non-representational, enactive knowledge acquired in these interactions is the basis of intercorporeality and empathy. Explicit or propositional forms of knowing others (“theory of mind”) are derived from later steps of development; they are not sufficient for explaining the interactive and empathic human capacities. This will finally be illustrated by the example of autism.

## 1 Introduction

Gilbert Ryle’s seminal contraposition of two fundamental forms of knowledge, *knowing that* and *knowing how* (Ryle 1949), may be traced back to Aristotle’s distinction in the ‘Nicomachean Ethics’ between *epistēmē* and *technē*, the first meaning theoretic or scientific knowledge, the latter skill or craft. Bertrand Russell (1910) has proposed the terms “*knowledge by acquaintance*” and “*knowledge by description*”, pointing out that the former is obtained through a direct interaction with objects or situations, whereas the latter is acquired in an indirect way, namely based on propositional language, for example, through description or explanation.

There has been a long and still ongoing debate on whether these two types of knowledge belong to distinct categories, or whether one depends upon the other, and if so, which is the more basic one.<sup>1</sup> However, if it is true that in our

---

<sup>1</sup> Ryle himself as well as Dreyfus (1972) argue that *knowing how* as a realization of skills, i.e. complex dispositions, may not be translated into propositional statements, whereas Stanley (2011) takes the opposite intellectualist position: “Knowing how to do something is the same as knowing a fact” (l.c., vii).



intuitive dealings with situations we always “know more than we can tell” (Polanyi 1967), and if we thus could never fully describe how to dance a Tango or to recognize the “mischievous” expression in someone’s face, then this implicit knowing may in principle not be completely converted into declarative or symbol-based knowledge.

With regard to developmental psychology too, it quickly becomes clear that our everyday relation to the world is based on a fundamental familiarity and skilfulness which is already acquired in early infancy before the development of symbolically and verbally mediated knowledge. At the end of their first year of life, babies are capable of highly differentiated interactions with persons and objects which doubtlessly fulfil the criteria of knowledge by acquaintance, long before conceptual knowledge of others is acquired. Here too, *knowing how* precedes *knowing that* (on this, see below).

Nevertheless, concepts of knowledge that may be described as propositional, symbol-based or representational are clearly prevalent in present-day cognitive and social sciences. Whether the question is how we recognize the world in general or how we perceive the social other, the way the question is posed already assumes a distant observer who learns about the world not through practical interaction, engagement and participation but through detached description, modelling or reconstruction. To acquire knowledge then means to form an idea, a representation or a model of the object or of the other, on which basis one can then proceed to action. Accordingly the mind is conceived as a system of inner models, constructs or representations which today are localized in the brain and enable the prediction of changes in the external world. In this way, however, the knower and the known, or cognition and the world remain separated from each other on principle.

Two related approaches are currently challenging this cognitivist paradigm, namely the *embodied and enactive approach to cognition* on the one hand and the *phenomenology of the lived body* on the other. Both seek to overcome the dualism of representational mind and external world by regarding conscious experience as a person’s being in the world through the medium of the body. According to the enactive paradigm, perception and action are inherently connected (Varela et al. 1990, Thompson 2007): Feeling a surface is accomplished through the *act of touching*, seeing an object is enabled through the *activity of looking*, etc. Moreover, each perception already evokes possibilities for action, that means, the objects are accessible for us, “ready-to-hand”, in Heidegger’s terminology, offering affordances for our mobile body (Gibson 1979). In these perception-action cycles, however, inside and outside, or mind and world can no longer be separated.

In the same way, phenomenology regards consciousness not as a self-enclosed entity, but as “being-towards-the-world” through the medium of the body (Merleau-Ponty 1962), or in other words, as the intentional and practical relation of the embodied subject to the objects and situations it finds itself in. Embodiment is the primordial form of subjectivity, but it is at the same time spatiality, situatedness, directedness to a horizon of possibilities which offer themselves to the body. On this assumption, however, the basic presupposition of representationalism has to be dropped. For representations ‘stand for something’ of which they must be separated. Now if the world is constituted for us only in the ongoing interaction with it, and if we are always already bodily acting in the world, then there is no separate “inner” which could map, reconstruct or represent the “outer”. In a constant circular process, no segment can stand “for another”. This does not exclude representations *within* conscious experience – for example, memories, imaginations, ideas of absent objects – but defies the monadic conception of consciousness itself as an internal representation or modelling of the world.

This has consequences for the question which kind of knowledge is more basic – knowing-that or knowing-how. If the objects are primarily given or ready-to-hand through our embodied interactions with them, and if the world is thus always already disclosed through the medium of the body, then representational knowledge appears only later on the scene. It is called for in the very moment when the ongoing, pre-reflective interaction with the world or with others is disturbed or interrupted. To use Heidegger’s example, when a hammer breaks, it loses its usefulness and appears as merely *there*, “present-at-hand”, and becomes a problem to be solved. Similarly, when the primary, empathic, interbodily communication with others suffers an irritation or disturbance, then we become aware of them as “beings of their own” which are not really transparent for us. In such situations of rupture we start reflecting or theorizing about the objects or the others, asking for explanations, causes, mechanisms, or in other words, seeking *knowledge about* them instead of relying on *knowledge by acquaintance*. This irritation and the attempt to overcome the resulting worry may be regarded as the root of *epistêmê* or science in general.

The distinction between knowing-that and knowing-how has gained additional momentum through the investigation of *implicit memory*. This kind of memory is formed in the course of the interaction of organism and environment: Recurring patterns of interaction have sedimented in the form of sensorimotor schemas and corresponding bodily dispositions. We may speak of an implicit “body memory” that underlies our everyday habits and skills, without necessarily being made explicit (Fuchs 2000, 2012). This embodied knowledge is realized in suitable situations through habitual action or through overarching volitional

acts. It then connects body and environment through ongoing cycles of perception and action that are based on earlier experiences.

In what follows, I will first present the concept of embodied knowledge, then connect it with the notion of body memory in order to elaborate its developmental dimension. I will then further elucidate the structure of embodied knowledge by taking the example of acquiring *social skills* through dyadic interactions in early childhood. I will argue that the non-representational, enactive knowledge acquired in these interactions is the basis of intercorporeality and empathy. Explicit or propositional forms of knowledge about others (“theory of mind”) are derived from later steps of development which presuppose the capacity of perspective-taking. They are not sufficient, however, for explaining the basic empathic human capacities, or the knowing-how of intercorporeality.

## 2 Embodied knowledge

Now what is embodied knowledge? – According to Ryle, *knowing that* is information-based knowledge that can be asked for and communicated in propositional language. In contrast, *knowing how* refers to training-based knowledge (e. g., how to ride a bicycle, how to dance a waltz) that cannot be reduced to a set of propositions. It consists of dispositions for integrative perceptions and actions which are enacted by the body without targeted attention. Granted, there is no strict separation between both knowledge systems. Knowing that and knowing how together may contribute to intelligent behaviour, as can be seen in the case of an experienced surgeon (Ryle, 1949, p. 49). Propositional knowledge then provides higher-level strategies which in turn are realized through practical, embodied know-how. Thus, in skilful coping, top-down and bottom-up approaches work together and influence each other reciprocally.

The traditional cognitivist approach, however, has no concept of knowing how; instead, it conceives of the mind as a disembodied system of representations and predictive models that are separated from embodied action. The standard information-processing model has a “sense-think-act” structure (Pfeifer & Scheier 1999; Pfeifer & Bongard 2007): First, the mind is supposed to *represent* the situation on the basis of stimuli processing (“sense”), then it computes the suitable behaviour (“think”), and finally it issues the corresponding command for bodily motion (“act”). Thus, there are three clearly divided stages of cognition and action: input, inner computational process, and output.

This model disrupts the unity of mind and body interacting with the environment in ongoing feedback cycles. In embodied action, there is neither place nor time for a separate goal representation which could then produce the necessary

movement. Instead, bodily skills and environmental affordances work together in a moment-to-moment process of continuous adjustment and fine-tuning. There is no hidden mind that directs the body based on deliberations or calculations. In his *Phenomenology of Perception* (1945/2012), Merleau-Ponty describes the example of knowing how to typewrite as follows:

[O]ne can know how to type without knowing how to indicate where on the keyboard the letters that compose the words are located. Knowing how to type, then, is not the same as knowing the location of each letter on the keyboard, nor even having acquired a conditioned reflex for each letter that is triggered upon seeing it. [...] It is a question of a *knowledge in our hands*, which is only given through a bodily effort and cannot be translated by an objective designation. The subject knows where the letters are on the keyboard just as we know where one of our limbs is – a *knowledge of familiarity* that does not provide us with a position in objective space. (Merleau-Ponty, 1945/2012, p. 145, emphasis added)

This knowledge is not representational, but it is not subpersonal either, that means, only to be found in neurally realized cognitive systems. It is enacted by the lived or subjective body, the body that I am myself as the agent of typing. “Consciousness is originally not an ‘I think that’, but rather an ‘I can’” (p. 139), as Merleau-Ponty writes. “I can” does not mean the conscious control of bodily motions, but rather a prereflective, protentional awareness of possible movements that accompanies each action. It is spread over the body, as it were, on the basis of the sensorimotor body schema. Through its habits and skills, the body anticipates or rather *implies* potential actions: It is prone to act in a way that is influenced both by its acquired dispositions and by the affordances or possibilities of the present situation. The more skilled and habitualized the body’s action, the less we are conscious of it, as William James has put it: “Consciousness deserts all processes where it can no longer be of use.” (James 1950, p. 496) In short, the body is the subject which knows how to act.

### 3 Implicit and explicit knowledge

Embodied knowledge may also be conceived, in Polanyi’s terms, as *tacit* or *implicit knowledge* (Polanyi 1967) in that it cannot be explained or verbalized explicitly. When our body parts coordinate while dancing a tango, when we perceive the expression of anger in someone’s face, or when an experienced psychiatrist intuitively makes a diagnosis on the basis of various symptoms and his overall impression of the patient, there is each time more knowledge involved than we can tell. A major reason for this is that the forms of knowing how are based on intermodal and sensorimotor *gestalt units*, that means, they inte-

grate different sense modalities and bodily movements into a holistic experience, as becomes obvious in tango dancing – think of the typical swing and rhythm of a movement and its musical grounding. In contrast, verbal articulation may only explicate single strands out of this undetermined-manifold clew of holistic experience. Thus, it is able to class these strands into a general context and to render them available to communication – yet at the price of losing the immediacy and unity of intuitive experience.

Our primary experience consists of holistic impressions, encompassing gestalts of perception and movement, whereas the single elements are only explicated secondarily: “It is the explication of the implicit which first generates delimitable components of meaning, namely through creating precision, selective emphasis and semantic relations.” (Jung 2014, p. 76; my translation). Only poetic language is able to evoke holistic impressions. It achieves this through rhythmicity and prosody, multi-valent usage of words and last not least through the self-referential frame of the artwork: A poem does not refer denotatively to a pre-given reality as does ordinary speech. Of course, the impression thus evoked is again an experience that may not be fully transformed into propositional language.

The explicating analysis of implicit experiences into single elements generally runs the risk of losing the primary phenomenon. Examples for this abound: A perceived facial expression is lost if one pays attention to its single features or details. Similarly, if we focus on a body part, it often no longer functions as a component of implicit capacities. A musician who pays attention to his individual fingers during a passage will easily make a mistake, and a tango dancer will look ridiculous once he moves his legs deliberately like a beginner. Practitioners in many skilled movement domains are aware that self-conscious thought can disrupt well-practised actions.

Interestingly, a *pathological loss* of embodied knowledge may be found in schizophrenic patients who often experience a fragmentation of holistic perception into single details:

I have to put things together in my head. If I look at my watch I see the watch, watchstrap, face, hands, and so on, then I have got to put them together to get it into one piece. (Chapman 1966)

A schizophrenic patient of Minkowski’s was no longer able to read because ‘[...] he became attached to a word, a letter, and did not attend to the meaning of the sentence. He examined whether all the “I”s had dots over them, whether there were accents where needed, whether all the letters had the same form.’ (Minkowski & Targowla 2001, p. 273)

Similarly, the units of habitual action sequences may dissolve, resulting in a pathological explication and hyperreflexive awareness of normally tacit aspects of everyday behaviour:

If I want to do something like going for a drink of water, I've to go over each detail – find cup, walk over, turn tap, fill cup, turn tap off, drink it. (Chapman 1966, p. 239).

At times, I could do nothing without thinking about it. I could not perform any movement without having to think how I would do it. (de Haan & Fuchs 2010)

These pathological cases illustrate again that the implicit structure of embodied knowledge conveys a holistic mode of existence which cannot be replaced by explicit reconstruction. The body acts as the medium of our relation to the world precisely inasmuch as it withdraws into the background of awareness. It conceals itself precisely in the act of revealing the world (Leder 1992). Turning our attention backwards on our embodied skills and habits tends to dissolve the spatial and temporal gestalt units on which they are based.

## 4 Body memory

Implicit knowledge or knowing how is not just an innate property of the body, but develops and constantly changes over the whole life-span. The acquisition of skills and habits has come to be explored in cognitive psychology under the heading of “*implicit*” or “*procedural memory*” (Schacter 1987, 1999), for which I will use the more encompassing phenomenological notion of *body memory* (Fuchs 2008, 2012).

Body memory may be defined as the entirety of established practices and skills that are available through the medium of the lived body without the need to remember earlier situations. Habits formed through repetition and practice are activated of their own accord; well-rehearsed sequences of movements have been incorporated, thus becoming a bodily capacity – like the upright gait, speaking or writing, using instruments like a bicycle, a typewriter or a piano. This bodily memory, which was first considered by Maine de Biran (1799/1953) and Henri Bergson (1896/2007), does not re-present or “presentify” the past, but rather re-enacts it in the ongoing conduct of life. In the last analysis, all capacities during one’s life point to a primordial capacity of the embodied subject, to a basic “I can” (Husserl, 1952, 253).

There are two major ways of acquiring bodily habits and skills: On the one hand, we can *explicitly synthesize* single elements of perception and movement through deliberate training. What is perceived or performed piece by piece at

first is gradually integrated and incorporated as a novel skill. Thus, we have learnt at primary school to spell and connect single letters until we could read the whole words and sentences. We have learnt to dance a tango by combining the single movements until the body had integrated them into an overarching flow of rhythm, dynamics and movement. Granted, these learning processes are based on pre-existing gestalt units (the word as heard, the swing of bodily movement, etc.) in which the explicit elements may be integrated.

On the other hand, many skills and habits are acquired *implicitly or unnoticed*, namely as a “learning by doing”, just through repeated practice, be it in dealing with objects or through interacting with others. Indeed, the most fundamental skills which have disclosed the world for us and upon which our everyday practices are based have sedimented into our body memory in the first and second year of life without any explicit teaching. This applies in particular to a type of memory to which we owe the skill of bodily interacting with others, and which I call *intercorporeal memory* (Fuchs 2012). In what follows, I will look at some stages of its development.

## 5 Intercorporeal memory

Infant research has shown that newborns are already able to imitate facial expressions of others like frowning, opening of the mouth, protruding the tongue, etc. (Meltzoff & Moore 1977, 1989), later on also emotional expressions such as smiling or surprise. This shows that they are equipped with an innate body schema which enables them to translate the seen gestures into their own felt movement, thus gaining a basic sense of familiarity with others. As early as in the first months, infants are also capable of discerning emotions such as happiness, sadness, and surprise in the postures, movements, facial expressions and vocal intonations of others (Hobson 2005, pp. 39). The basis for this is that different sense modalities and movements can have the same ‘kinematics’ and thus express the same affect, which may best be rendered in *musical* qualities (‘crescendo’, ‘decrecendo’, flowing, bursting, pulsing, etc.). The feeling of joy and the various expressions of joy have similar intermodal dynamics, and this is the basis for the direct perception of others’ emotional states even in earliest childhood.

Affectivity, however, is primarily not an inner or individual state, but a dyadic experience of mother and infant, mediated through expression and bodily resonance in subtle gestural, facial and vocal interactions. Already 6–8 weeks after birth, so-called “proto-conversations” arise, that means, alternating vocalizations and gestures (Trevarthen 1979), overall a fine-tuned co-ordination of movements and expressive signals which may be compared to a couple dance. These phenom-

ena of “interbodily resonance” (Fuchs u. De Jaegher 2009) and “affect attunement” (Stern 1985) generate encompassing emotional states: The emerging affect during a joyful playing situation between mother and infant may not be divided and distributed among them, but arises from the shared social situation. Emotions are primarily embedded in intercorporeality and interaffectivity.

Moreover, recurrent patterns of interaction and affect attunement are sedimented as interactive schemas in the infant’s body memory (“*schemes of being-with*”, Stern 1985), for example “mummy-feeding-me”, “daddy-playing-with-me”, etc. This results in what Stern (1998) has called *implicit relational knowing* – an embodied, intuitive knowledge of how to interact with others, how to have fun together, how to elicit attention, to avoid rejection etc. It is a temporally organised, “musical” memory for the rhythm, dynamics and undertones which resonate in the interaction with others. Thus, long before verbal communication infants already acquire a primary understanding of others through shared practices recorded in their intercorporeal memory.

## 6 Embodied empathy and its disturbance in autism

This is the basis of a primary form of empathy that emerges in face-to-face encounters: In embodied interaction, the other is not assumed ‘behind’ his action, but he enacts and expresses his feelings and intentions in his conduct. Embodied relational knowledge conveys an intuitive awareness of the other’s affective state. In perceiving his expressive movements and actions as embedded in the shared context, “[...] one already sees their meaning. No inference to a hidden set of mental states is necessary.” (Gallagher & Zahavi 2008, 185) Moreover, in social interaction, one’s own body is affected by the other in various forms of bodily resonance, sensations, tensions, movement tendencies etc. This resonance forms part of the embodied knowledge that allows us to understand the other’s state.

In contrast to this account, the currently predominant theories of social cognition are mainly based on representationalist approaches: Concepts such as Theory of Mind, mentalization or mindreading (Antonietti 2006, Goldman 2012) assume a fundamental inaccessibility of the other whose hidden mental states, intentions or feelings may only be inferred from his external bodily behaviour by using some sort of rule-based ‘mindreading’. Social cognition would thus be based on observation, inference and *knowing that*, even though it may not always be expressed in propositional terms.



However, our primary and everyday encounters with others are not observations from a 3<sup>rd</sup> person point of view, but embodied interactions within the 2<sup>nd</sup>-person perspective. In these, we normally don't use any imaginative modelling or inference; instead, we immediately perceive the other's intentions and emotions in his expressive behaviour. Of course, we may sometimes apply methods of explicit conjecturing or inferring another's mental state. This happens in particular when an irritation, misunderstanding or disturbance occurs, and we ask ourselves why the other said or did what he did, what he might be thinking or feeling, etc. We can then also transpose ourselves into the other, take his perspective, reason about his motives, search for his hidden intentions etc. As I argued at the beginning, *knowing that* is called for when the ongoing, pre-reflective interaction with the world or with others is disturbed. But it is not the primary or default mode of social understanding, and it is only acquired much later on in childhood. Rather, implicit intercorporeal or relational knowing forms the basis of intersubjectivity. This may finally be illustrated by another psychopathological example, namely *autism*.

As is well known, children with autism lack the basic emotional contact with others, which leads to a variety of social, communicative and cognitive deficits. According to current cognitive theories of autism, the disorder is due to a difficulty to "read other people's minds," or to imagine what they are thinking or feeling. The suggestion is that autistic people lack a "Theory of Mind" (ToM) – the purported neural or cognitive device that computes others' underlying intentions from their perceived behaviours. In recent years, however, criticism has been raised by phenomenological psychiatrists and philosophers, arguing that the deficit is rather caused by failures of early interaction and intercorporeality (Hobson 1993, Gallagher 2004, Fuchs 2015). This is supported by the fact that many autistic symptoms such as lack of emotional contact, of interest in others, agitation and anxiety are already present in the first years of life, that means, long before the supposed age to acquire a ToM which is around 4 years.

I have pointed out before that knowing how is essentially based on processes of gestalt formation that enable us to perceive and act in a holistic way instead of being aware of the single elements. This applies for social cognition as well: The expression of a face is only perceived when we do not focus on a single feature or detail. Now it has been demonstrated that autistic children show problems precisely in establishing *perceptual and situational coherence*: They focus on single parts or elements rather than perceiving the gestalt of objects, and they tend to treat things and events decontextualized, thus missing their particular meaning provided by the situation as a whole (Frith 1989, Happé 1995).

Accordingly, eye tracking studies have shown that children with autism focus on peripheral features of faces, and on irrelevant details of interactive sit-

uations while missing the relevant social cues (Klin et al. 2003). This failure of holistic cognition may have some positive effects such as remembering unrelated or non-sensical items, however, it significantly interferes with the development of social understanding. Thus, as I have pointed out above, affect attunement is crucially based on perceiving emotional cues (gestures, facial movements, voicings) as holistic expressions and as embedded in recurrent situations. Similarly, understanding the intentions of others depends on learning how to relate their gestures and actions to the context in order to grasp their meaning.

Such deficits of autistic children converge to a fundamental disturbance of embodied social perception and interaction very early in life. They are not able to acquire the implicit relational knowledge that is based on schemes of being-with-others taken up into one's intercorporeal memory. What autistic children lack is thus not a theoretical concept of other minds but a primary sense of bodily being-with-others. ToM-like strategies of explicit mentalizing and inferring from social cues are rather employed by high-functioning autistic individuals as a *compensation* for the lacking capacities of primary intercorporeality. Thus, Temple Grandin, a woman with Asperger's syndrome, described her problems with interpersonal relations to Oliver Sacks as follows:

It has to do, she has inferred, with an implicit knowledge of social conventions and codes, of cultural presuppositions of every sort. This implicit knowledge, which every normal person accumulates and generates throughout life on the basis of experience and encounters with others, Temple seems to be largely devoid of. Lacking it, she has instead to 'compute' others' intentions and states of mind, to try to make algorithmic, explicit, what for the rest of us is second nature. (Sacks 1995, p. 270)

These compensatory strategies enable functional interactions with others to a certain degree, but fail to establish the primary sense of being-with-others which is normally conveyed by intercorporeality and implicit relational knowledge:

She is now aware of the existence of these social signals. She can infer them, she says, but she herself cannot perceive them, cannot participate in this magical communication directly [...]. Knowing this intellectually, she does her best to compensate, bringing immense intellectual effort and computational power to bear on matters that others understand with unthinking ease. This is why she often feels excluded, an alien. (Sacks 1995, p. 272)

As we can see from Grandin's report, the implicit relational knowledge mediated by the body and its intercorporeal memory cannot be substituted by explicit inference or rule-based theorizing; in other words, knowledge by acquaintance *with others* may not be replaced by knowledge *about them*.

## 7 Summary

Embodied knowledge is the foundation of our familiarity with the world and with other people. It is a knowledge and skill which is realized in perceiving and reacting on situations, without needing targeted attention or memory. The subject of knowing is itself embodied: It finds these knowings and skills not inside, but only in its practical engagement with the world. In contrast, representational, symbol-based forms of knowledge – knowing that – arise from an indirect, secondary relation to the world which the human mind is capable of by taking a distance from objects and situation and representing them as such. This presupposes, however, that the world is already disclosed to us via the medium of the body which has acquainted itself with the world from birth on. We experience the world, because our body has become *transparent* for it; that means, we experience the implicit actions and affections of our body *as* the objects and situations of our environment.

Following Polanyi (1967, 1969) we may describe this transparent structure of our experience as an interplay between the “distal” pole, i.e. the thematic, explicit or focal object of awareness, and the “proximal” or bodily pole, which recedes from attention and is known only in a tacit, non-thematic manner:

Our body is the only assembly of things known almost exclusively by relying on our awareness of them for attending to something else [...] Every time we make sense of the world, we rely on our tacit knowledge of impacts made by the world on our body and the complex responses of our body on these impacts. (Polanyi 1969, p. 147)

The body is thus “passed over in silence”, as Sartre (1956) put it. Inasmuch as we perceive or act through an organ of our body, “it necessarily recedes from the perceptual field it discloses” (Leder 1990, p. 14), and the same applies to the skills that are realized by the organs. Thus as a medium, the body withdraws in the tacit dimension; “it conceals itself precisely in the act of revealing what is Other” (Leder 1990, p. 22), and yet remains the core of our self. The transparency of the body arises precisely from the embodied nature of the mind.

Therefore in the application of embodied knowledge or skills we are not dealing with a blind or even subpersonal occurrence which we could only ascribe to a (neuro-)physiological process. The skills that are based on body memory, such as tango dancing, are realized by tuning in to the familiar rhythm of movement and enacting its particular style. It is I myself who is dancing, not a body machine commanded by a disembodied mind. This allows me to conduct and modulate the body’s enactments, like a conductor its orchestra, without having to generate them, for they happen of their own accord.

As we can see, embodied knowledge unburdens our attention from an abundance of details, thus facilitating our everyday performances. The body and the senses become a medium through which the world is accessible and available. We are capable of directing our attention toward the gestalt and the meaning of what we encounter. Action is facilitated, as we may intend its goal instead of noticing every single movement. The will becomes free since the bodily means and components of acting recede into the background. A primary goal-directed intention suffices to release the complete arc of action. While his fingers move the keys, the pianist is able to direct himself to the music itself, to listen to his own play. Thus, freedom and art are essentially based on the tacit knowing how of the body.

The body's familiarity with the world is not innate, however, but is based on a primary disclosure of the world which, using a stoic notion, may be termed *oikeiosis* (from the Greek *oikos* = house, home), that means, "indwelling" or becoming acquainted with one's home. *Oikeiosis* develops in early childhood in the course of embodied interactions with the world and with others, as described above. In these, infants feel perceived and accepted by their caregivers and, embedded in this affective resonance, they can acquire the skills of dealing with objects and situations. The disclosure of the world happens primarily through *knowledge by acquaintance* in shared, intercorporeal practices. Thus, familiarity with the world and with others are equiprimordial and inseparable foundations of the lifeworld. As soon as representational or symbolic forms of cognition and knowledge develop, they permit an extended understanding of the world and open up new possibilities for action. However, they remain always dependent on the primordial familiarity with the world which the body had already established before we became aware of it.

## Bibliography

- Beebe, Beatrice/Lachmann, Frank/Jaffe, Joseph (1997): "Mother-infant interaction. Structures and presymbolic self- and object representations". In: *Journal of Relational Perspectives* 7, pp. 133–182.
- Bergson, Henri (2007) : *Matière et mémoire. Essai sur la relation du corps à l'esprit*. Paris: PUF.
- Russell, Bertrand (1910): "Knowledge by acquaintance and knowledge by description". In: *Proceedings of the Aristotelian Society* 11, pp. 108–128.
- Chapman, James (1966): "The early symptoms of schizophrenia". In: *British Journal of Psychiatry* 112, pp. 225–251.
- De Biran, Maine (1953): *Influence de l'habitude sur la faculté de penser* [The influence of habits on the faculty of thinking; originally published 1799]. Paris: PUF.

- de Haan, Sanneke/Fuchs, Thomas (2010): "The ghost in the machine: Disembodiment in schizophrenia. Two case studies". In: *Psychopathology* 43, pp. 327–333.
- Dreyfus, Hubert (1972): *What computers can't do: The limits of artificial intelligence*. New York: Harper & Row.
- Frith, Uta (1989): *Autism: Explaining the enigma*. Oxford: Basil Blackwell.
- Fuchs, Thomas (2000): "Das Gedächtnis des Leibes". In: *Phänomenologische Forschungen* 5, pp. 71–89.
- Fuchs, Thomas (2008): "Das Gedächtnis des Leibes". In: T. Fuchs, *Leib und Lebenswelt. Neue philosophisch-psychiatrische Essays*. Kusterdingen: Die Graue Edition, pp. 37–64.
- Fuchs, Thomas (2011): "Psychopathologie der Hyperreflexivität". In: *Deutsche Zeitschrift für Philosophie* 59, pp. 565–576.
- Fuchs, Thomas (2012): "The phenomenology of body memory". In: S. C. Koch/T. Fuchs/M. Summa/C. Müller (eds.): *Body memory, metaphor and movement*. Amsterdam: John Benjamins, pp. 9–22.
- Fuchs, Thomas (2015): "Pathologies of intersubjectivity in autism and schizophrenia". In: *Journal of Consciousness Studies* 22, pp. 191–214.
- Fuchs, Thomas/De Jaegher, Hanne (2009): "Enactive intersubjectivity: Participatory sense-making and mutual incorporation". In: *Phenomenology and the Cognitive Sciences* 8, pp. 465–486.
- Gallagher, Shaun (2004): "Understanding interpersonal problems in autism: interaction theory as an alternative to Theory of Mind". In: *Philosophy, Psychiatry & Psychology* 11, pp. 199–217.
- Gibson, James J. (1979): *The ecological approach to visual perception*. Houghton Mifflin, Boston.
- Happé, Francesca (1995): *Autism: An introduction to psychological theory*. Cambridge: Harvard University Press.
- Hobson, Peter (1993): *Autism and the development of mind*. Hillsdale NJ: Lawrence Erlbaum Associates.
- Hobson, Peter (2005): *The cradle of thought*. Oxford: Oxford University Press.
- Husserl, Edmund (1952): *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. II. Phänomenologische Untersuchungen zur Konstitution*. Husserliana Vol. IV. Den Haag: Martinus Nijhoff.
- James, William (1950): *The Principles of psychology* (1890). New York: Dover Publications.
- James, William (1890): *The principles of psychology*. Volume One. New York: Henry Holt and Company.
- Jung, Matthias (2014): *Gewöhnliche Erfahrung*. Tübingen: Mohr-Siebeck.
- Klin, Ami/Jones, Warren/Schultz, Robert/Volkmar, Fred (2003): "The enactive mind, or from actions to cognition: lessons from autism". In: *Phil. Trans. R. Soc. Lond.* 358: pp. 345–360.
- Leder, Drew (1992): *The absent body*. Chicago: University of Chicago Press.
- McGhie, Andrew/Chapman, James (1961): "Disorders of attention and perception in early schizophrenia". In: *British Journal of Medical Psychology* 34, pp. 103–116.
- Meltzoff, Andrew/Moore, M. K. (1977): "Imitation of facial and manual gestures by human neonates". In: *Science* 198, pp. 74–78.
- Meltzoff, Andrew/Moore, M. K. (1989): "Imitation in newborn infants: exploring the range of gestures imitated and the underlying mechanisms". In: *Developmental Psychology* 25, pp. 954–962.

- Merleau-Ponty, Maurice (1962): *Phenomenology of perception*. Transl. C. Smith. London: Routledge and Kegan Paul.
- Minkowski, Eugène, Targowla, R. (2001): “A contribution to the study of autism: The interrogative attitude”. In: *Philosophy, Psychiatry, and Psychology* 8, pp. 271–78.
- Pfeifer, Rolf/Bongard, Josh (2007): *How the body shapes the way we think: A new view of intelligence*. Cambridge, Mass.: MIT Press.
- Pfeifer, Rolf/Scheier, Christian (1999): *Understanding intelligence*. Cambridge, Mass.: MIT Press.
- Polanyi, Michael (1967): *The tacit dimension*. Garden City, N.Y.: Anchor Books.
- Polanyi, Michael (1969): *Knowing and being*. Ed. M. Grene. Chicago: University of Chicago Press.
- Russell, Bertrand (1910–1911): “Knowledge by acquaintance and knowledge by description”, In: *Proceedings of the Aristotelian society (new series)*, Vol.XI, pp.108–128
- Sartre, Jean-Paul (1956): *Being and nothingness*. New York: Philosophical Library.
- Schacter, C. L. (1987): “Implicit memory: history and current status”. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13, pp. 501–518.
- Schacter, Daniel L. (1999): *Wir sind Erinnerung. Gedächtnis und Persönlichkeit*. Rowohlt: Reinbek.
- Stanley, Jason (2011): *Know how*. New York: Oxford University Press.
- Stern, Daniel (1985): *The interpersonal world of the infant*. New York: Basic Books.
- Stern, Daniel (1998): “The process of therapeutic change involving implicit knowledge: Some implications of developmental observations for adult psychotherapy”. In: *Infant Mental Health Journal* 19, pp. 300–308.
- Thompson, Evan (2007): *Mind in life. Biology, phenomenology, and the sciences of mind*. Cambridge/Mass: Harvard University Press.
- Trevarthen, Colwyn (1979): “Communication and cooperation in early infancy: a description of primary intersubjectivity”. In: Margaret Bullowa (Ed.): *Before speech*. Cambridge: Cambridge University Press, pp. 321–347.
- Varela, Francisco/Thompson, Evan/Rosch, Eleanor (1991): *The embodied mind: Cognitive science and human experience*. Cambridge, MA: MIT Press.



Michel Bitbol

## Panpsychism in the First Person

(In the existentialist branch of phenomenology), writes a critic, one has turned off the mind. It's just the opposite. It has been put everywhere, because we are not mind and body, we are not a consciousness facing the world, but embodied mind, being-in-the-world  
– Maurice Merleau-Ponty

**Abstract:** A central presupposition of science is that objectivity is universal. This does not only create a blind spot in knowledge, but also forces one to ignore it. Several strategies were accordingly adopted in to overcome this ignorance, along with the standard divide between continental and analytic philosophy. One of them is *Phenomenology*, with its project of stripping the layers of interpretation by way of a complete suspension of judgment (*epochè*), and evaluating any claim of knowledge from such a basis of “pure consciousness”. Another one is *pan-experientialist metaphysics*, that puts back pure experience in the very domain that was deprived of it by the act of objectification. I compare these two approaches, thereby establishing a hierarchy of radicality between avoiding the blind spot from the outset and compensating for it retrospectively.

### 1 Introduction

Pure experience is elusive because it is not. It is lying at the permanent blind spot of what there is, for the mere reason that it constitutes the precondition of anything that *is*, namely of anything that may *appear*. Conversely, sharpening perceptual differences, improving the efficiency of technology, increasing the discrimination of phenomena by scientific theories, perpetuates the blind spot of living and knowing. Something, which being no *thing* is all the more easy to forget, remains in the dead angle of knowledge; and no move forward can *ever*, in *principle*, account for what has been left behind by this very push.

This impossibility clause expresses a limitation of science. But science also has a more positive teaching in store about the issue of lived experience. Indeed, scientific advances progressively squeeze the domain of their own silent lived precondition, thereby avoiding incorrect characterizations. The background premise of perceiving, reasoning, and knowing is found to be no entelechy, no “*élan vital*”, no ghost-like soul, and no spirit; for the task of those speculative entities has been carefully identified by scientific research and ascribed to objective “mechanisms”. Even the words “mind” and “consciousness”, which were supposed to capture the residual enigma left in knowledge by objectification, have been redefined in terms of



blind cognitive functions. Mind is understood as a system of information processing that allows problem solving and decision making. As for consciousness, it is torn between its original meaning which includes lived experience, and a more abstract meaning that only encompasses the functions of meta-cognition and synthesis of representations (Block et al. 1999, p. 375).

To sum up, scientific advances are correlated with a lexical retreat. What escapes scientific characterization becomes elusive, because any name which may be given to it, is confiscated and endowed with a cognitive meaning. With no name and no method of handling, “this” is considered (Cohen & Dennett 2011) as “outside the scope of science” (which is correct), and thereby negligible (which is disputable).

The stakes of the search of a name for the elusive background are then high; not, of course, to elaborate a self-contradicting objective science of the pre-objective, but to avoid neglecting the latter. To begin with, the word “consciousness” may have unwanted connotations. Bertrand Russell pointed out that consciousness is supposed to be *intentional* (to be consciousness *of* some object), and that it can hardly encompass non-intentional states like pleasure or anxiety. Moreover, consciousness is usually associated with reflectivity, whereas what we are trying to pick out comes prior to any reflection. “Pure experience”, borrowed from William James (1976, p. 117), is a reasonable alternative. Indeed, experience accommodates non-directional tonalities, and it is said to be “pure” in view of its anticipating any conceptual structure or self-realization. So much so that it has been found suitable by the Japanese philosopher and founder of the Kyôtô school, Nishida Kitarô (1990, p. 3), as a valid expression of the deconstructive way of Zen. Yet, the word “experience” (as any other word that purports to refer to what is referring) also has its drawbacks. It suggests that there is a *subject* of experience, be it a “thin”, non-permanent, subject (Strawson 2006, p. 191). The subject-object polarity is still roaming around.

An alternative lexicon has then been suggested by authors who lean towards panpsychism but are reluctant to ascribe elementary entities of the material world the elaborated synthetic, reflective, and self-centered consciousness which is realized in human beings. Accordingly, their alternative vocabulary relies on mass terms rather than on count terms. Whitehead, after Peirce, offered the word “feeling” (Whitehead 1929, p. 236). William Clifford also used this term to designate a kind of subject-free experience (“*sentitur* is all that can be said”); but immediately after, he noticed that a feeling is usually a *complex*, and that we still need a name for its “element” (Clifford 1879, p. 84–85). “Mind stuff” was accordingly retained by Clifford as a denomination of the simple, plain, subject-free “feel” of which the universe is said to be made. Much later, even the mentalistic undertones of the panpsychist vocabulary was found misleading,

and names for some sort of “proto-mental” *continuum* were sought. Herbert Feigl (1960) chose the neutral word “quality”, and Ken Wilber (2000) adopted the fascinating word “depth”, to capture the elusive proto-*psychè*. Even Nishida Kitarô (1958) abandoned the Jamesian expression “*pure experience*” in favor of a more idiosyncratic “*place of absolute nothingness*”, which aims at denoting a process of self-hollowing out that gives room to manifested beings.

At this point, it is clear that:

1. Science is leaving a gaping hole, which is nothing less than the most glaring evidence.
2. Science prevents one from conceiving this hole as some occult entity or property that is formally similar to objective entities and properties, yet (mysteriously) inaccessible to its methods.

If we wish to obtain an exhaustive framework in which to accommodate the seeing as well as the seen, the manifestation as well as the manifest, some dramatic initiative must then be taken.

This kind of initiative can develop along two different directions that were mixed up until now. These two directions were taken, respectively, by analytic philosophy and by continental philosophy. The first approach is tantamount to moving in the same “progressive” direction as scientific research, yet speculating in order to identify a missing property of things. The second approach consists in adopting the opposite direction of “regression” towards the “realm of the mothers of knowledge” (Husserl 1976, p. 174), namely towards the experiential background out of which the world of manifest objects is picked out and constituted. I will document both approaches in succession, starting with a critical exposition of the “progressive” one, which identifies with panpsychist metaphysics, and then defending the “regressive” one, as the most promising in terms of making sense of our being-in-a-world.

## 2 Panpsychism as an *a posteriori* filling in of the blind spot of science

Panpsychist metaphysics starts from a naturalist assumption and two crucial arguments (Skrbina 2005, p. 250). The assumption is that there is only one reality, which is *physical*, and optionally that physical reality consists of interacting particles of matter (Nagel 1979). The two key arguments are: (1) “the phenomenon whose existence is more certain than the existence of anything else (is) experience” (Strawson 2006), and (2) there can be no “emergence” of experience, in

view of its radical heterogeneity to any standard physical property. The first argument is usually considered as very powerful. It was expressed as a compelling *deduction* by Descartes (“I think therefore I am”, or “I am thinking therefore I exist.” (Descartes 1985, p. 127)) But these Cartesian sentences are deceitful in so far as their syllogistic form is utterly restrictive. Behind their logical pattern, there is an immediate intuitive certainty (Nishida 1990) popping out from the obvious impossibility of *living* their negation. Even Dennett, the archetypal opponent of the thesis of certainty, has retreated from (or at least nuanced) his initial assertion that conscious experience is somehow an illusion (Blackmore 2002). In a metaphysical context, this chief argument turns out to have momentous ontological consequences. Bertrand Russell has expressed it beautifully, in terms that are reminiscent of William Clifford’s construal of mental stuff as a basic constituent of the universe: “Everything that we know of (the world’s) *intrinsic* character is derived from the mental side” (Russell 1992, p. 402). In other terms, we can access the world as it is in itself only by the direct acquaintance offered by lived experience (Griffin 1997); any other kind of knowledge is derivative and subject to doubt. The same conviction was later conveyed by Herbert Feigl (1960), the main proponent of the mind-brain identity theory, who added that, by contrast with this first-hand, direct, intrinsic, access to what there is, physical science is only concerned by indirect access to observables features and structural knowledge.

Let’s come now to the no-emergence argument. This one is more difficult to shield against criticism, but it is reasonably sound. It has gained some credit from Galen Strawson’s description of what emergence of unexpected macroscopic features out of the interaction of known microscopic features would require. According to him, “[t]here must be something about X and X alone *in virtue of which* Y emerges, and which is *sufficient* for Y” (Strawson 2006). The emergence of (say) liquidity from an interaction of water molecules is relatively easy to account for, by way of their Van der Waals interactions; but it looks like there is nothing in the physico-chemical description of atoms and molecules “in virtue of which” phenomenal consciousness should arise. This obstacle was obliquely confirmed by a chief proponent of the emergentist thesis: Jaegwon Kim. In his book entitled *Physicalism or something near enough*, the expression “near enough” expresses a challenge to a standard emergentist view of phenomenal consciousness. Indeed, after a reflection in which Kim documents accepted cases of emergent features as “*weakly emergent*” (their emergent status being only *apparent*, due to limitations of our conceptual or perceptive abilities), he concludes that phenomenal consciousness is a *unique* case of *strong* emergence, since it is necessarily unpredictable from the microscopic properties of physical elements (Kim 2005, 2006). *Strong* emergence, here, can be read as a combination of postulated emergence *plus* irreducible ignorance

about the “in virtue of which” this should take place. If one is not dogmatic about the postulate, ignorance is what is left.

In this gloomy situation (for emergentism), the panpsychist strategy looks inviting. It amounts to broadening our concept of the physical by adding to it a new kind of property. This new kind of property purports to be commensurable to what is missing in the physical picture of the world (phenomenal consciousness), yet simple enough to avoid ascribing full-blown conscious minds to every minute physical entity or aggregate of such entities. Since the “psychè” of panpsychism usually encompasses too many advanced features of the human mind, such as reflectivity, self-consciousness, mental processing, complex emotions etc., this word has been replaced by more fundamental terms. The most uncommitted word is “*protopsychism*”, giving rise to the doctrinal denomination “pan-protopsychism”. Another, more precise, one is “experience”, giving rise to “panexperientialism” (Nixon 2010).

But this is only the verbal aspect of a systematic rush towards elementarity. The rush started with Leibniz’ “little perceptions” (Leibniz 1993) of simple monads, which are compared with mathematical differentials, and opposed to the integral “apperception” of high-level monads. The same rush is still developing currently. Attempts at characterizing the elementary spark of phenomenal consciousness have been made by using the standard correspondence between mentality and behavior (Lewtas 2013). The bottom-level experience is then characterized as follows: devoid of structure and intelligence, non-representational, fleeting, transient, repetitive, first-order (as opposed to the second-order thoughts of reflectivity), patchy, homogeneous, yet *sentient*. A (non-trivial) task is to understand how this raw material can be added up, or “integrated”, into a full consciousness: this is the well-known “combination problem” (Seager 1995), which remains a stumbling block of panpsychism.

This being granted, one perceives a strong tension spread across the network of panpsychist doctrines. On the one hand, the basis and starting point of this family of doctrines is first-person experience, and on the other hand, the stance of this (as of any other) speculative metaphysics is objectifying, external, spectator-like. Even though its concept can be stretched beyond recognition, nature is understood by the culturally dominant view as a horizon of description for objective science, whereas sentience is a concept isolated from the direct acquaintance of our own pre-objective experience. Panpsychism here looks like a baroque combination of idealism and physicalism, which turns out not to be viable in the long run.

The dilemmas of panpsychism are in fact quite acute. To begin with, the choice of taking first-person experience as a starting point is not consistently pursued. At first sight, the sought consistency is achieved by the demand of an ontological continuity in nature (Whitehead 1967) in terms of elementary experiences instead of

physical properties. In the same way as the materialist requires that the manifest universe be made of physical entities and properties up to the highest level of organization, the panpsychist claims that it (also) consists in “feelings” or elementary experiences down to the lowest level of organization. But actually, this is a pretence, because unlike in materialism, the ontological continuity of panpsychism does not go together with existential continuity. The ontological continuity reached by materialism arises from constant adoption of the existential stance of intentional directedness and search for objectification of phenomena. True, a massive feature (the experiential *non*-thing) is forgotten or neglected as a consequence of this goal-oriented stance of materialism, but this is the price to be paid for methodological unity and efficiency. By contrast, the ontological continuity of panpsychism (experience all the way down the ladder of an objectified nature) relies on adopting *two* incompatible stances: (i) the receptive and reflective stance needed to realize that the basis of inquiry is indeed lived experience, and (ii) the intentional directedness needed to pinpoint the elements of nature to which experiential properties are ascribed.

A host of difficulties derives from this existential stretching which underpins panpsychism. As mentioned previously, many panpsychists claim to be physicalists, or at least objectivists. But they are bound not to endorse all the consequences of a picture of the objective world derived from physics. This half-way position was beautifully and synthetically expressed by Galen Strawson: a panpsychist is likely to be a physicalist, but not a *physicS-alist*. In other terms, despite her overt commitment to philosophical physicalism, a panpsychist philosopher cannot restrict her ontology to the observable objective properties of actual physics, since she is precisely trying to complete her representation of nature with “feelings”, “depths” or elementary experiences.

One consequence of this unstable position is that physics is only used as a vague inspiration for various plausible narratives about the objective world. Let me give two examples. A mereologically inclined panpsychist may take advantage of a mereologic view of objectified nature to prop up her doctrine. She may establish a one-one correspondence between her atoms of sentience and the elementary particles of high-energy physics, trying to figure out what it is like to be such a particle (Lewtas 2013). By contrast, a holistically inclined variety of panpsychist is likely to prefer the field-theoretic representation of quantum physics (or even classical electromagnetism), thus connecting the power of synthesis of consciousness with the continuum of a field. These two varieties of panpsychists then hold antipodal representations of the world, borrowed from various domains of physics chosen for their analogy to a given function of consciousness. But they also share a common mistake. Both of them seem to ignore that no such representation is to be taken at face value, as a picture of nature-as-it-is-in-itself, but rather as a picture-*like* tool for orienting predictions and actions (Bitbol 1996;

Van Fraassen 2010). In other terms, the majority of panpsychist physicalists adopt alternately a naïve attitude towards physics, by adhering to the dubious ontological claims of its popularizers, and a distant attitude towards the sound epistemological lessons that can be drawn from it.

Another unwelcome consequence of the panpsychists' combination of physicalism and distance with respect to physics, is that they must define the "physical" *in abstracto*. A widespread proposition of this sort is borrowed from Neurath's well-known characterization of physicalism (Neurath 1983): this doctrine is concerned by phenomena *located and extended in space-time*. Panpsychism is then defined as the position that holds that (i) "the universe is spatio-temporal in its fundamental nature" (Strawson 2006), and (ii) spatio-temporally located elements are endowed with protopsychical qualities. Space-time here seems to be given the status of a repository of those "external" entities and qualities whose "interiority" is protopsychical. This may be acceptable from the standpoint of an intentional/objectivistic stance, but not from the standpoint of the receptive/reflective stance. In the latter stance, even the spatial oppositions of internal and external, right or left, higher or lower, are supposed to arise from the structure (Ruyer 2012) and/or activity of the knower. This conflict of two opposite conceptions of space is no novelty. It was already vivid at the beginning of the eighteenth century, when orthodox Newtonians defended an absolutistic version of space, whereas Leibniz considered that space expresses the system of relations of (sentient) monads. And the same conflict was brought to completion by Kant, who contrasted a transcendental realist and a transcendental idealist conception of space-time as "*a priori* forms of sensibility". Thus, even the minimal, *non-physicalist*, expression of physicalism is at odds with the allegedly experiential starting point of panpsychism. As Erwin Schrödinger cogently pointed out, once lived experience has been left aside in order to elaborate an objective picture of the world, "[i]f one tries to put it in or on, as a child puts colour on his uncoloured painting copies, it will not fit. For anything that is made to enter this world model willy-nilly takes the form of scientific assertion of facts; and as such it becomes wrong" (Schrödinger 1986, p. 148)

To recapitulate, panpsychists have so much focused on recovering an ontological continuity in nature without forgetting its experiential feature, that they have shattered their own existential continuity in doing so, thus generating a host of difficulties and conceptual conflicts. Borrowing bits and pieces from physicalism and attempting to bring them to completion by adding a first-person-like ingredient turns out to be a deadlock.

### 3 Phenomenology as *a priori* fullness

From this point on, the way is open to a completely different strategy that would not only *accommodate* the glaring fact of lived experience, but that would stick to it consistently throughout. This strategy is phenomenology, which is a descendent of *transcendental idealism*. But the alternative strategy is no caricatural or metaphysical version of idealism. Instead of claiming that consciousness is all there is, it contents itself with advocating a receptive/reflective stance and making sense of everything else (including science) from the standpoint of that stance. Husserl's phenomenology is no theoretical idealism, but rather an attempt at describing faithfully the experiential soil of practical life (Husserl 2007, p. 48). This strategy is no doctrinal assertion about the *nature of Being* (say ideal, as opposed to spatio-temporal), but the uncompromising adoption of a *way of being*.

Let us further realize at this point that the epistemological lessons of modern physics (especially quantum physics) are perfectly consonant with the basic stance of phenomenology. Indeed, from the standpoint of a reflective stance, it is natural to understand theories in terms of the mental and performative conditions for acting *in* the world, rather than in terms of what they capture of the *nature of* an allegedly independent world. Along with this approach, physical theories are not concerned with *representing* the world as it is in itself, but rather with expressing the structure of a wide class of adaptive transformations. Kant thus showed that the structural core of Newtonian mechanics can be derived from the general preconditions of objective knowledge of spatio-temporal objects called material bodies (Kant 2004). Far from having an empirical origin and a representational status, this structure is provided in advance by the project of objectifying a fraction of what is presented to us in sensible experience. A similar strategy can be adopted to understand reflectively, rather than representational-ly, the structure of quantum theory (Bitbol 1998, 2011).

Adopting the phenomenological stance then has momentous consequences for the *premises* of the problem of consciousness. Firstly, the starting point of inquiry in experience is not only confirmed by phenomenology but also stabilized and explored. Secondly, from a phenomenological standpoint, no reason is left for sanctifying the “physical” of physicalism. For a phenomenologist, if “physical” is taken to mean “observable in space-time”, adopting a physicalist doctrine is tantamount to turn incorrectly the spatio-temporal objects of perceptive attention and intentional directedness into intrinsically existing beings. And, if “physical” means falling into the domain of validity of physics, or being encompassed by the concepts of a physical theory, it becomes even more obvious to a phenomenologist that no ontological status can be granted to it. Indeed, from a phenom-



enological standpoint, those concepts are nothing more than by-products of an enactive (Varela et al. 1992) process of sense making applied to what appears. The basic slogan of panpsychism is thus completely overturned. According to panpsychists, *there is more to the physical (namely to nature) than what a physics of observable objective properties can tell us* ; and only in this “more” can the nucleus of lived evidence be accommodated. By contrast, along with a phenomenological approach, the said “more” does not come after, but *before* any observation and any theorizing has occurred. Antecedently to what appears, there is appearance; antecedently to things that can be shown, there is the showing; antecedently to present items, there is presence.

Many difficulties of panpsychism are automatically dispelled by adopting the phenomenological stance. Firstly, phenomenology is not underpinned by a couple of mutually exclusive positions or attitudes (the reflective and the objectifying) ; it rather understands the need and meaning of each approach, including the objectifying one, from the standpoint of its receptive/reflective attitude taken as a reference. Secondly, phenomenology is not concerned with the “combination problem” of panpsychism, namely the problem of building a whole of synthetic experience out of a plurality of elements of experience dispersed in the world. It bypasses this problem not, of course, because it relies on a holistic physical representation, but because it bases its enquiry on a *gestalt*-like starting point: the present, global, embodied experience of the researcher. The typical question of phenomenology is not “how does the whole of experience arises from protoexperiential sparks?”, but rather “how can one describe by means of discontinuous words a present integrated experience?”. Thirdly, the method of phenomenology for accessing the most elementary forms of experience is neither rational nor observational, but lived. Unlike panpsychism, phenomenology does not rely on the “behavior” of physical objects, to figure out indirectly what it might be like to have a simple transient experience. Instead, it uses a direct method of access that digs through our cumulative kind of experience, stripping it layer after layer of its perceptive or intellectual fabrications, and trying to reach beneath it a deeper layer of pure presence. This method is called the *epochè* (suspension of judgment and attentional focusing), and it is assisted by successive “reductions” of what appears to *strata* of less and less elaborated structures of appearance. The method can be radicalized by meditation, culminating with the ability to see *any* present object or mental activity as a naked presence (this is the realization of “suchness”, *Tathatā* in Sanskrit). Fourthly, the ontological issue is not skipped by phenomenology but rather transformed beyond recognition. Instead of adopting a distantiated ontology of observable objects (as in the natural sciences or a panpsychist extension of this distantiated ontology (that encompasses *conceivable* atoms of experience), phenomenology adumbrates an ontology of immersion, of connivance, of acquaintance. Phenomenology, in its



mature form, looks for an “oblique ontology” of intertwining (Saint Aubert 2006), or, as Merleau-Ponty writes, an “endo-ontology” (Merleau-Ponty 1964, p. 279). This is an ontology expressed *from* the innermost recesses of the process of being, rather than an ontology *of* the external contemplation of beings.

A (may be too) simple theoretical tool offered by Husserl to make sense of the regressive path that goes from elaborated consciousness to elementary experience, is the distinctions he made between: the (sensory) matter and the (pre-conceptual) form of experience; the *hylè* and the *morphè*; the noesis (flux of lived experience) and the noema (structure of the intentional object). One may then oppose: (i) the non-intentional hyletic and noetic components of consciousness that represent what is left after experience has been stripped of its interpretative layers, and (ii) the noematic intentional structures that have to be bracketed, neutralized in order to make the former component appear. But this immemorial separation between matter and form is undermined when Husserl points out that even the interpretative layer made of intentional acts and noematic structures, partakes of what is the case in pure experience: “What is lived includes in its real (*reell*) composition not only the hyletic moments (colours, sensory sounds), but also the apprehensions which animate them” (Husserl 1950, p. 339).

Michel Henry took advantage of the latter reflection to disrupt Husserl’s hylemorphism. According to him (Henry 1990), any experience, be it the experience of an “external” object, or the experience of abstract forms, or the experience of one’s own mental acts, ultimately is an experience of the *self-affection of one’s flesh*. Russell’s examples of pleasure and pain are taken as paradigmatic, and even intentionality, even the assumption of transcendence of natural objects, must be rooted in the immanent impression that the flesh of the living being is making on itself. In other terms, even the *morphè* is experienced as *hylè*; even the perception of a patch of colour on some “outer” object is underpinned by a self-perception of the perceiver. Intentional consciousness is borne by non-intentional experience, and therefore the deepest layer of the archaeological stratification of consciousness is nothing else than a naked self-sensitivity.

*Original* can only mean this: what comes in itself before any intentionality and independently of it, before the space of a gaze, before the “outside oneself” of which intentionality is only a name; what comes [...] before the world, out of the world, foreign to any conceivable ‘world’, *a-cosmic*. (Henry 2000, p. 82)

In other terms, what comes before intentionality, and before the belief in a world, is the non-directional impression of being embodied, the silent voice of the body whose psychological name is “cenesthesia”. But once again, unlike the naturalistic program, which would account for the latter impression in terms of natural

objects in the human body, the phenomenological program adopts a diametrically opposite stance. It starts from the deepest layer of what we are, and then justifies the belief in a natural world as a consequence of the differentiation of such primeval experience.

This phenomenological program was radicalized in Merleau-Ponty's posthumous book, *The Visible and the Invisible*. According to Merleau-Ponty, "we can accept a world ... only after having witnessed its arising from our experience of raw Being, which is like the umbilical cord of our knowledge, and our source of meaning" (Merleau-Ponty 1964, p. 209). The first move must be to identify the characteristics of the "experience of raw Being". In order to do so, Merleau-Ponty does not rely on discursive thought, but on tacit "thought-experiments", which are in fact *experiential* experiments. Let me give two explicit examples of those experiential experiments, as close as possible to everyday life. The first one is to be performed by opening our eyes and weakening progressively the tension towards identifying a variety of objects in the visible environment. If we do that, we cease to see ourselves as a bodily object in front of other objects, and we realize something even more stunning than Ernst Mach's well-known "self-portrait". We perceive distinctly that the seeing emerges from the midst of what is seen, from a sort of hole of transparency which breaks through its dense impenetrability, and that this density is boundless, extending far beyond our skin. The second experiential experiment is done with our eyes shut. It consists in rubbing a cloth or a smooth stone, concentrating the attention on the sensation that arises from this contact, and suspending judgments about objects. Then, what is felt is an undifferentiated mass of sensation in an atmosphere of reversibility: one can hardly say if what is felt is the friction of the hand on the cloth, or the caress of the cloth on the hand, for there is no contrast between the two feelings.

The discursive expression of what is experienced in such occurrences could be called radical *situatedness*. Indeed, in view of this kind of experiential experiments, we are not onlookers of a nature given out there, but rather intimately intermingled with it. We are not point-like spectators of what is manifest, but a field of experiences which merges with what appears in a certain region of it. From a theoretical standpoint, this immediately gives rise to what has been called an *endo-ontology*, an ontology of the participant *in* being, rather than an ontology of the observer *of* beings. Here, Being is not presented before me as an object of sight, but my vision arises from the middle of Being. "Vision is the tool which allows me to be absent to myself, and to contemplate from within the fission of Being." (Merleau-Ponty 1985)

From an experiential standpoint, the way is open to realizing what is usually neglected due to educational prejudice, namely our true complicity with a nature which has too hastily been isolated from us, as if it could be reduced to an object

of perception and thought. But as soon as we have settled in this state of coincidence, what do we perceive exactly, or, more exactly, what is perceived through us? Merleau-Ponty answers this question with a single word: our *flesh*. We perceive our flesh not as something separate, but as a self-perceived perceiver. The flesh is that strange being endowed with reversibility, since it is jointly perceived and self-perceiving. The archetypal case of a two-faced kind of perception is the sense of touch, which, unlike distanced vision, is synchronously appearance of what is touched and self-revelation of the touching in its carnal thickness. Besides, the sense of touch has also a less organical way of revealing the twofold nature of the flesh, in so far as one can perceive by means of one part of one's own body (say the left hand) the softness and resistance of some other part (say the right hand). Here, one is no longer dealing with the fusion of standard perception and intrinsic self-perception, but with a duality of status and functions (*toucher* and *touched*) realized by one and the same body. This second, milder, case of reversibility is also realized by vision, since the flesh is both *seen* as this (human) visible body, and *seeing* from a certain hollow location of this body in the middle of the head.

This description of our own situation is almost trivial. But dazing triviality is the subject matter of phenomenology. Phenomenology is this unique discipline that sharpens our sensitivity to trivial yet primeval facts of experience which are usually ignored due to our exclusive fascination for objects.

From this point on, Merleau-Ponty turned triviality into a thrilling epic of being. He invited us to let go any ready-made objectified analysis of the world, especially its division into bodies (mine and yours, living and inert etc.). He recommended to become fully receptive to experience as it is, with its unbounded, massive, opaque style of appearing, instead of perceiving it through the glasses of our adult educated system of distinctions. As soon as this is done, the outcome is stunning. Even the identification of the flesh to *this* precise body is bound to melt, since the process of self-definition and positing boundaries between bodies has been voluntarily suspended. The whole surrounding world is then perceived as an extension of "my" flesh, and it thereby becomes *the flesh of the world, the world as flesh*. "Where should we locate the boundary between the body and the world", Merleau-Ponty asks, "since the world is flesh?" (Merleau-Ponty 1964, p. 182) There are no such things as "me", "you", and the world, but a single canvas wherefrom various self-individualizing centers of sensitivity emerge, and which leaves patches of elementarity and half-obscurity between these centers (Barbaras 1993, p. 304). The role of constituting objectivity, which had been entrusted to the transcendental *ego* by Husserl is further extended to whatever has the status of a flesh by Merleau-Ponty. But since the flesh is boundless, any division between the constituter of objectivity and the constitut-

ed objects is meaningless. Just as the flesh is self-perceiving, the world *qua* flesh is self-objectifying.

## 4 Conclusion

We can see at this point some similarities and a huge gap between panpsychism and Merleau-Ponty's boundless phenomenology of the world-flesh. In both cases, any difference of nature between mental and mindless, sentient and inert beings is denied. In both cases, experience suffuses what there is. This point of convergence is so deep and so striking that one is surprised to find a recent historian of panpsychism declaring that there is no equivalent of a panpsychist tendency to be seen anywhere in phenomenology (Skrbina 2005). But it is also true that panpsychist approaches and Merleau-Ponty's phenomenology use so deeply antagonistic methodologies, that the former similarity can easily be overridden. Panpsychist metaphysics starts its inquiry with the centers of meaning of an adult activity and objective knowledge; and it sticks on them sparkles of experience whose content is imagined by a behaviorist extrapolation. In other terms, panpsychism takes the ordered *cosmos* of reason and science for granted, and painfully adds to it what looks absent in it. By contrast, phenomenology is a strategy of extreme deconstruction of the *cosmos* of reason and science, not because it ignores it but because it purports to reach its fountainhead. Such fountainhead can be exhumed below the strata of perceptive and scientific interpretations, and it turns out to be an unelaborate embodied experience of self-sensitivity. Then, as soon as this embodied experience is brought to (its own) light, it appears that there is, and can be, nothing else. The usual distinction between experience and its worldly objects relaxes, and what remains is only a world of experience. To sum up, the difference is between gluing bits of experience onto an allegedly given "external" world, and (re)discovering the world *qua* experience, neither external nor internal.

## Bibliography

- Barbaras, Renaud (1993): *De l'être du phénomène*. Paris: Jérôme Millon.
- Bitbol, Michel (1996): *Schrödinger's philosophy of quantum mechanics*. Dordrecht: Kluwer.
- Bitbol, Michel (1998): "Some steps towards a transcendental deduction of quantum mechanics". In: *Philosophia Naturalis* 35, pp. 253–280.
- Bitbol, Michel (2011): "Reflective metaphysics : understanding quantum mechanics from a Kantian standpoint". In: *Philosophica* 83, pp. 53–83.

- Blackmore, Susan (2002): "The grand illusion: why consciousness exists only when you look for it". In: *New Scientist*, June 22/2002, pp. 26–29
- Block, Ned/Flanagan, Owen/Güzeldere, Guven (eds.) (1999): *The nature of consciousness*, Cambridge: MIT Press.
- Clifford, William K. (1879): *Lectures and essays*. Vol. II. London & New York: Mac Millan.
- Cohen, Michael/Dennett, Daniel (2011): "Consciousness cannot be separated from function". In: *Trends in Cognitive Science* 15, pp. 358–364.
- Descartes, René (1985): "Discourse on the method IV". In: *The philosophical writings of René Descartes*, volume 1, Cambridge: Cambridge University Press.
- Feigl, Herbert (1960): "The mind-body problem: not a pseudo-problem". In: Sydney Hook (ed.): *Dimensions of mind*. New York: New York University Press.
- Griffin, David (1997): "Panexperientialist physicalism and the mind-body problem". In: *Journal of Consciousness Studies* 4, pp. 248–268.
- Henry, Michel (1990): *Phénoménologie matérielle*. Paris: Presses Universitaires de France.
- Henry, Michel (2000): *Incarnation*. Paris: Seuil.
- Husserl, Edmund (1950): *Idées directrices pour une phénoménologie*. Paris: Gallimard.
- Husserl, Edmund (1976). *La crise des sciences européennes et la phénoménologie transcendantale*. Paris: Gallimard.
- Husserl, Edmund (2007). *De la réduction phénoménologique*. Grenoble: Jérôme Millon.
- James, William (1976). *Essays in radical empiricism*. Cambridge MA: Harvard University Press.
- Kant, Immanuel (2004). *Metaphysical foundations of natural science*. Cambridge: Cambridge University Press.
- Kim, Jaegwon (2005). *Physicalism or something near enough*. Princeton: Princeton University Press.
- Kim, Jaegwon (2006). *Philosophy of mind*. Boulder: Westview Press
- Leibniz, Gottfried Wilhelm (1993). *Nouveaux essais sur l'entendement humain*. Paris: Flammarion.
- Lewtas, Pat (2013). "What it is like to be a quark". In: *Journal of consciousness studies*, 20, pp. 30–64.
- Merleau-Ponty, Maurice (1964). *Le visible et l'invisible*. Paris: Gallimard.
- Merleau-Ponty; Maurice (1985). *L'œil et l'esprit*. Paris: Gallimard.
- Nagel, Thomas (1979). *Mortal questions*. Cambridge: Cambridge University Press.
- Neurath, Otto (1983). "Physicalism: the philosophy of the Vienna Circle". In Robert Cohen and Marie Neurath (eds.). *Philosophical papers 1913–1946*. Dordrecht: Reidel.
- Nishida, Kitarō (1958). *Intelligibility and the philosophy of nothingness: three philosophical essays*. Honolulu: East-West Center Press.
- Nishida, Kitarō (1990). *An inquiry into the good*. Yale: Yale University Press.
- Nixon, Gregory (2010). "From panexperientialism to conscious experience: the continuum of experience". In: *Journal of Consciousness Exploration & Research*, 1, pp. 213–401.
- Russell, Bertrand (1921). *The analysis of mind*. London: George Allen & Unwin.
- Russell, Bertrand (1992). *The analysis of matter*. London: Routledge.
- Ruyer, Raymond (2012). *Néo-finalisme*. Paris: Presses Universitaires de France.
- Saint Aubert, Emmanuel de (2006). *Vers une ontologie indirecte : sources et enjeux critiques de l'appel à l'ontologie chez Merleau-Ponty*. Paris: Vrin.
- Seager, William (1995). "Consciousness, information, and panpsychism". In: *Journal of Consciousness Studies* 2, pp. 272–288.

- Skrbina, David (2005). *Panpsychism in the West*. Cambridge MA: MIT Press.
- Strawson, Galen (2006). *Consciousness and its place in nature*. Exeter: Imprint Academics.
- Van Fraassen, Bas (2010). *Scientific representation: paradoxes of perspective*. Oxford: Oxford University Press.
- Varela, Francisco/Thompson, Evan/ Rosch, Eleanor (1992). *The embodied mind*. Cambridge MA: MIT Press.
- Whitehead, Alfred North (1929). *Process and reality*. Cambridge: Cambridge University Press.
- Whitehead, Alfred North (1967). *Science and the modern world*. New York: The Free Press.
- Wilber, Ken (2000). *Integral psychology*. Boulder: Shambala.



Karl Mertens

# What Is It Like to Be an Angel?

## The Importance of Corporeally-Situated Experience for Our Concept of Action

**Abstract:** This article seeks to show that any sound theory of action needs to take the bodily and situated nature of self-experience into account. To this aim, Nagel's famous question 'what is it like to be a bat?' is modified by asking: 'What is it like to be an angel, i.e., a creature which is not bodily situated in the world?'. In order to explain what is entailed by the idea of the disembodied existence of angels, the considerations refer to what Thomas Aquinas treats as the essential features of angels' existence. Following Nagel, it is subsequently argued that the reasons why we cannot imagine what it is like to be an angel from an angel's perspective are immediately clear and indisputable. However, there still remains another version of Nagel's question: what is it like to be an angel from a human being's perspective? It is argued that, for reasons related to the constitutive situatedness of experience, this question cannot be answered either. In contrast to Nagel's considerations about bats, the present article is not interested in defending an epistemological claim on the subjective quality of consciousness against the claim of reductionism. Showing that we cannot possibly conceive of what could be called a human angel's world, the article rather develops some tentative and programmatic considerations about what the limitations of our imagination imply for our self-understanding. Emphasizing such limitations allows us to defend the need to appropriately account for the structures of bodily situated existence in order to develop an adequate theory of action.

In his famous article "What Is It Like to Be a Bat?" Thomas Nagel designs an intriguing thought experiment (Nagel 1974). He requests us to imagine the quality of the experiences of a bat; i.e., what it is like to be an animal whose behavior is based on an orientation through sonar. Even if we readily concede that the bat's orientation is a possible kind of sense perception, we cannot imagine the concrete quality of such a perception because its structure differs fundamentally from our own experiences. As a consequence, although we ascribe awareness and sensitive perceiving to bats and might even be able to imagine what it would mean for us to be a creature like a bat, we cannot imagine what it is like to be a bat from the bat's own perspective. In the final analysis, this experiment fails, thereby demonstrating the limits of our imagination. For us, it is impossible to imagine the specific quality of experiences of creatures whose per-



ceptual structures essentially differ from ours. However, it does not fail because a bat's experience is not possible. On the contrary, Nagel's thought experiment is firmly based on the idea that bats have sensory experiences. What these reflections show is that any inquiry into consciousness needs to account for the inseparable relationship between forms of consciousness and a species-related subjective perspective.

In the following, I would like to modify Nagel's thought experiment by asking a similar question: What is it like to be an angel? – It doesn't matter if you happen not to believe in the existence of angels. All we need is a concept of what may be called an angel. For this purpose, I refer to some classical remarks by an expert in philosophical questions concerning angels: Thomas Aquinas. However, unlike Aquinas, I am interested in taking a phenomenological stance toward angels by asking whether we are able to describe a kind of experience which can possibly be ascribed to angels. If we follow Nagel, it becomes immediately clear that we cannot imagine what it is like to be an angel from an angel's perspective. However, there remains another version of Nagel's question I would like to deal with: What is it like to be an angel from a human being's perspective (cf. Nagel 1974, p. 169)? In contrast to Nagel's considerations about bats, I am not interested in an epistemological thesis that defends the subjective quality of consciousness against the claim of reductionism. Showing that we cannot possibly conceive of what could be called a human angel's world, I will instead present some tentative and programmatic considerations about what the limitations of our imagination imply for our self-understanding. I especially want to stress the importance of our bodily situated existence for an adequate theory of action.

## 1

To start, I would like to single out the most striking essential characteristic of those beings called angels: In contrast to us, angels are not corporeal entities; they have no bodily existence. This widespread view about angels also appears in Thomas Aquinas, who states: “*Angeli non habent corpora sibi naturaliter unita*” / ‘angels do not have bodies naturally united to them’ (S.th. I, q. 51, a. 1, c.).<sup>1</sup> Aquinas positively characterizes an angel as ‘*creatura pure spiritualis*’, a “purely spiritual creature” (S.th. I, q. 50). In contrast to creatures like us and animals that are – to take up a Platonic metaphor – imprisoned in a fading mor-

---

<sup>1</sup> Slight modifications of the Latin text (for grammatical reasons) or partial revisions of the English text are marked by simple quotation marks.

tal body, angels are imperishable, i. e., they are immortal (S.th. I, q. 50, a. 5, c.). As purely spiritual beings, they have intellect and will. According to Aquinas, these potencies do not require corporeal organs to realize their activities (cf. S.th. I, q. 54, a. 5, c.).

At first glance, it is already clear that angels are not situated in the world like human beings. I would like to elaborate on this a little bit more. Since angels are not localized in a body, they have no sensible perception. Although Aquinas argues that an angel is aware of material things, according to him this awareness is not sensory, but occurs rather through a pure intellectual power (cf., f.e., S.th. I, q. 57, a. 1, c.). Furthermore, angels have no affections because incorporeal entities have no sensitive part of the soul (S.th. I, q. 54, a. 5.). Moreover, it is problematic even to ask whether an angel perceives its environment, since an angel does not have a spatio-temporal ‘environment’ as we do. Angels do not exist in space and time – at least not in the sense we do.

Although Aquinas states that angels have a place and are able to move under the conditions of time (cf. S.th. I, q. 52, a.1, c.; q. 53, a. 3, c.) because they are created beings (cf. S.th. I, q. 52, a. 2, c.), this spatial and temporal existence has to be understood rightly. Regarding their spatial existence, angels do not have localized presence in the sense of tactile experience and extension. Rather, they are said to be virtually present wherever they happen to be acting. As Aquinas explains:

[A]n angel and a body are said to be in a place in quite a different sense. A body is said to be in a place in such a way that it is applied to such place according to the contact of dimensive quantity; but there is no such quantity in the angels, for theirs is a virtual one. Consequently an angel is said to be in a corporeal place by application of the angelic power in any manner whatever to any place. (S.th. I, q. 52, a. 1, c)<sup>2</sup>

Concerning their temporal existence, Aquinas clarifies that we cannot measure angelic time according to our human temporal existence. Angelic time refers to the sequentiality of angelic acts, not to a temporal extension. In any case it “is not the same as the time which measures the movement of the heavens, and whereby all corporeal things are measured” (S.th. I, q. 53, a. 3., c.).<sup>3</sup> As a consequence, we, as finite beings localized in worldly space and time, have no

---

<sup>2</sup> “[A]equivoce tamen dicitur angelus esse in loco, et corpus. Corpus enim est in loco per hoc, quod applicator loco secundum contactum dimensivae quantitatis. Quae quidem in angelis non est; sed est in eis quantitas virtualis. Per applicationem igitur virtutis angelicae ad aliquem locum qualitercumque, dicitur angelus esse in loco corporeo.”

<sup>3</sup> “Sed istud tempus [...] non est idem cum tempore quod mensurat motum caeli, et quo mensurantur omnia corporalia [...]”

means to adequately understand and grasp what it is like to have an angelic power and to be at a place or to move under conditions of time in the sense of angelic beings. In any case, Aquinas's explanation demonstrates that we cannot imagine the existence of an angel in a way that is comparable to our own existence as beings situated in a spatio-temporal context.

Assuming this rough concept of an angel, I think we can maintain, like Nagel, that it is impossible for us to imagine what it is like to be an angel from an angel's perspective. However, we are confronted with a deeper problem of imagination with regard to angels, because we cannot even understand what it is like to be an angel from a human perspective. The key reason has already been identified: While bats and humans have in common a bodily existence (even if their embodied experiences are quite different), angels have no corporeal being at all. In this respect, we are absolutely unable to imagine what it would be like for us to be an angel.<sup>4</sup> In the following I would like to highlight four reasons why we cannot answer this question.

## 2

First, one might object that according to Aquinas, there are at least some respects in which angels are like human beings. After all they have intellectual capacities as we have; they are able to think and to acquire knowledge. However, what does "thinking" mean in the case of a purely spiritual being? If we take a further look at the *Summa theologiae*, we can see that Aquinas describes the cognition of angels by negating essential features of human knowledge. He explains that whereas 'the lower, i. e., human, intellects reach perfection in knowing truth by a kind of movement and discursive intellectual operation; that is to say, as they advance from one known thing to another' (S.th. I, q. 58, a. 3, c.),<sup>5</sup> the angel's knowledge is not discursive.<sup>6</sup> 'The angel does not understand by reasoning, nor by composing and dividing' (S.th. I, q. 58, a. 4, c.),<sup>7</sup> i. e., by composing a pred-

---

<sup>4</sup> Note that the question "what would follow if we were angels?" is quite different from the intention of Aquinas who is interested in analyzing the nature of angels in the context of his theological-philosophical reasoning.

<sup>5</sup> "[I]nferiores intellectus, scilicet hominum, per quendam motum et discursum intellectualis operationis perfectionem in cognitione veritatis adipiscuntur; dum scilicet ex uno cognito in aliud cognitum procedunt."

<sup>6</sup> For a more detailed analysis of this and the following points see Goris 2012, pp. 169.

<sup>7</sup> "[R]elinquitur quod angelus, sicut non intelligit ratiocinando, ita non intelligit componendo et dividendo."

icate with a subject. In fact, for Aquinas, that which we would need a proposition to articulate, angels grasp in a single glance all at once. Moreover, since “an angel understands not by composing and dividing, but by understanding what a thing is” (S.th. I, q. 58, a. 5, c.),<sup>8</sup> an angel is not susceptible, as we are, to failures in the intellectual process of forming false sentences or false syllogisms (for details, see S.th. I, q. 58, a. 5, c.). Obviously such an intellect is quite different from ours, most of all because its understanding occurs without temporally extended acts of reflecting and reasoning. Hence angels are able to gain a form of knowledge that represents all spiritual things at once (cf. S.th. I, q. 58, a. 2, c.).<sup>9</sup> They survey the field of cognition in a single moment (S.th. I, q. 58, a. 3, c.).

In contrast to the angelic intellectual life as Aquinas explains it, the human intellect is constitutively characterized by the temporal features of a particular kind of activity. This is the case because our own mental life is essentially connected with our bodily existence. For example, even if we are dealing with problems of formal logic or mathematics, we must constantly work to keep our focus on a given problem and avoid distraction. We can merely concentrate our thinking efforts to a limited number of problems and usually only to one problem at once. And we can fail in our reasoning. Because of the constant possibility of error, we often must double-check our work. In short: even mere intellectual acts are exhausting for human beings. A chess player loses roughly three pounds of weight during an intense six-hour game. Thinking is hard work – for us.

Perhaps there are also human kinds of timeless thinking accompanied by a reduced level of bodily consciousness. In this context one might mention intellectual experiences like intuitive knowledge of principles in a moment, pure contemplation or what might be called intellectual flow experiences. These forms of thinking seem to be effortless; the thought or idea does not arrive bit by bit in a reasoning process, but rather just happens to a person instantaneously. However, these are exceptional cases of thinking which could not be characterized without reference to our ordinary corporeally-situated forms of thinking. Moreover, if we do achieve such an intellectual experience, it is embedded in arduous, temporally-realized cognitive processes in which we work through problems that are then suddenly solved by the corresponding insight, flow, or surmounted

---

<sup>8</sup> “[A]ngelus non intelligit componendo et dividendo, sed intelligendo quod quid est.”

<sup>9</sup> One should note, however, that Aquinas insists on the limitations of angelic knowledge in comparison to divine knowledge. The knowledge of angels is a knowledge “per speciem” which has to be distinguished from God who knows everything “per essentiam” because he himself contains the whole intellectual reality. For God’s knowledge according to Aquinas cf. S.th. I, q. 14, aa. 1 – 16. The differences and the relationship between the knowledge of God and of the angels are discussed, e.g., by Goris 2012, pp. 150 and pp. 167.

by contemplation. It is a moment in which the character of reflecting changes from an arduous, temporally-structured act to an event occurring to someone. Therefore, it is not experienced as an intellectual activity in the proper sense of an activity structured by temporal conditions, but rather as a form of extraordinary passive happening. Taking these considerations into account, we can say that we might have an intellectual experience of moments of an intellectual flow. However, we have no experience of the quality of a permanent flow. We cannot imagine the style of incorporeal thinking that does not at all require intellectual efforts of concentrating and focusing attention on an intellectual problem and that does not fatigue the thinking person after some time. Such thinking is different in kind rather than in degree, because we experience flow essentially as an exception, framed against our normal corporeal experience.

### 3

Second, according to Thomas Aquinas, angels are willing entities like us (cf. S.th. I, q. 59, a. 1, c.). Willing is a kind of pursuit of something aligned with an appropriate act realizing, or at least trying to realize, what is strived for. More accurately, Aquinas characterizes the angel's will by its tendency towards the good, which is finally realized by the cognition of the divine essence (cf. S.th. I, q. 59, a. 1, c.). Because of its immateriality, an angel's will does not refer to exterior acts – which means that it does not cause changes of physical states in the world. Rather, the will of an incorporeal creature applies to interior acts, i.e., to spiritual states like cognitions (cf. S.th. I, q. 59, a. 1, c.). It might seem as though certain kinds of human willing correspond to this angelic willing. For example the will to know or the will to control one's own will in the sense of Harry Frankfurt's second order volitions. – But if we try to understand and even to imagine this concept of pure spiritual willing more accurately, we encounter serious difficulties. Willing requires, at minimum, that the willing agent strive to cause an inner or outer state by his or her inner or outer action. This means that someone who wills is involved in a situation which can be changed by his or her will in at least some aspects. Therefore a willing creature is necessarily confronted with an inner or outer state that he, she or it wants to modify or overcome. Therefore willing something includes as a necessary condition the formulation of tasks and problems the willing person intends to solve. However, it is not even possible to give an adequate description of what situation might prompt an angel to will something. The reason is that all humanly conceivable ideas include reference to a bodily existence located in space and time. According to Aquinas, whereas 'every body is determined by the 'here' and 'now' (S.th. I, q. 50.

a. 1, c.),<sup>10</sup> angels are not subject to temporal-spatial conditions since they are incorporeal and immortal entities. However, without this situational background, we cannot truly imagine the concept of an angelic will as a kind of (pure spiritual) willing by means of comparison with the phenomenon of human willing.

Taking the aforementioned reflections about angelic intellect and will into consideration, we have to deny that an angel is able to act in a sense that is humanly comprehensible. Hence, it is hard for us (or even impossible) to understand how, as Aquinas states, intellectual and volitional powers do not require bodily organs (cf. *S.th.* I, q. 54, a. 5, c.), but only the capacities of intellect and will. The use of this power cannot be understood as a form of action, in the sense of something that alters physical or mental states in the world. As mentioned above, the intellect of an angel cannot be conceptualized as an act in the latter sense; and the angelic willing is not the will to change an inner or outer situation to be described by referring to a perspective of temporally structured bodily existence. In contrast, understanding how an angel may act requires us to consider how angels could be linked to a bodily existence. This converges with Aquinas who thinks it is possible for angels to use a corporeal body temporarily in order to become visible and perform actions (*S.th.* I, q. 51, a. 2, c.). Thereby angels are able to perform “*opera vitae*”, acts of (mortal) life (*S.th.* I, q. 51, a. 3, c.; cf. a. 2, c.). But Aquinas adds that even in this case the bodily movement which an angel performs is not necessarily limited to the spatial-temporal conditions of human actions (cf. *S.th.* I, q. 53, a. 1, c.; a. 3, c.).

By means of negation, we may gain from these considerations a deeper self-understanding of our own willing and acting which necessarily requires embodied agents who are involved in situations. It is common to understand action as a willed or intended behavior. Since it is intentional or voluntary, an action differs from mere behavior in essential respects. For this reason, the standard theory of action has centered around intentional concepts like “will” and “intention”. Consequently the contemporary philosophical discussion has extensively treated the question of how an intention can cause a physical state in the world. However, there are still relatively few reflections about the contexts and situations we are confronted with when we act (cf. the report in Leist 2007). If we consider the situations and contexts of human agency, some of the most fundamental insights are connected with the fact that agents are embodied persons.<sup>11</sup> Focusing on the meaning of a bodily existence of agents, philosophy of action not only turns out

---

<sup>10</sup> “[O]mne corpus determinatur ad hic et nunc.”

<sup>11</sup> This point is highlighted by fundamental reflections coming from a phenomenology of living bodies (see, for instance, Merleau-Ponty).

to be more concrete but also to reflect essential conditions which have been fairly neglected in the mainstream of the previous theory of action.

## 4

Third, there is no individuation without bodily existence. – So far, I have been speaking about angels as if they were individual persons. But speaking more strictly, this seemingly harmless claim has to be reconsidered now. It was stated above that for an angel as an incorporeal entity there is no worldly situation at all – at least in a human sense. As a result, the concept of an incorporeal individual existence is also deprived of meaning. Peter Frederick Strawson has pointed out clearly that a strictly incorporeal entity cannot be understood as an individual. To gain a self-understanding as an individual, an angel or another entity without a corresponding body must at least refer to a bodily existence in its own past. Strawson writes:

The other, and less commonly noticed point, is that in order to retain his idea of himself as an individual, he must always think of himself as disembodied, as a former person. That is to say, he must contrive still to have the idea of himself as a member of a class or type of entities with whom, however, he is now debarred from entering into any of those transactions the past fact of which was the condition of his having any idea of himself at all. [...] In proportion as the memories fade, and this vicarious living palls, to that degree his concept of himself as an individual becomes attenuated. At the limit of attenuation there is, from the point of view of his survival as an individual, no difference between the continuance of experience and its cessation. (Strawson 2006, p. 116)

The continuation of the quote adds a nice twist: “Disembodied survival, on such terms as these, may well seem unattractive. No doubt it is for this reason that the orthodox have wisely insisted on the resurrection of the body.” (Strawson 2006, p. 116) Taking a look at Aquinas’s angelology once more we can notice that for him – in contrast to philosophers like Scotus – individual angels, by virtue of their immateriality, must be conceptualized as a species (*species infima*) (cf. S.th. I, q. 50, a. 4, c.). The ontological background of this claim is Aquinas’s understanding of matter as the ultimate principle of individuation of creatures. Thus, without matter there can be no individuality. Again it is to be stated that the concept of an ‘individual’ angel as *species infima* essentially differs from the concept of individuality we use to distinguish living entities and which is decisive for our experience of individuality.

To summarize: We have to deny the possibility of asking in the style of Nagel what it is like to be an angel – be it from an angel’s perspective, be it from a

human perspective. The ultimate reason for this is that Nagel's question "what is it like to be" aims to develop the subjective quality of the consciousness or experience of a creature as a finite entity characterized by its own standpoint and perspective anchored in an individual bodily existence. However, in the case of angels there is no sense available to us to talk about a perspective, an individual standpoint or experience, because angels are incorporeal and not creatures situated in the dimensions of space and time familiar to us. – Against this background one can formulate the basic but not trivial insight that actions are performed by individual agents with a view of their own perspectives.<sup>12</sup>

## 5

Fourth, angels as creatures without a temporally and spatially situated body are not social beings in any sense understandable to us. We cannot conceive of a form of communication that does not presuppose mutual perception and interaction. In this sense an angel seems to be damned to a kind of solipsistic existence. Strawson brings out this solipsistic character of incorporeal existence in his chapter about persons in *Individuals* by pointing out that "[...] the strictly disembodied individual is strictly solitary, and it must remain for him indeed an utterly empty, though not meaningless, speculation, as to whether there are any other members of his class" (Strawson 2006, p. 115). However, we could even ask whether the concept of solitariness meaningfully refers to anything in the case of angels, since for beings without any worldly and social situation, the category of solitude would be meaningless insofar as it presupposes the experience of a social world.

I should point out that Aquinas would reject Strawson's view that incorporeal beings would be necessarily solipsistic, since he explicitly states that angels recognize one another. Yet, interestingly, they do this not by a kind of social communication, but rather by what he calls "internal word" (*verbum internum*). Aquinas thus posits a specific mechanism for angelic communication:

And by the fact that the concept of the angelic mind is ordered to be made known to another by the will of the angel himself, the concept of one angel is made known to another;

---

<sup>12</sup> This point has been elaborated by Paul Ricoeur in his criticism of the contemporary standard theory of action. According to Ricoeur, the theory of action as developed by Anscombe and above all by Davidson does not take into account the relevance of the agent because of its prevailing interest in analyzing action in the context of an ontology of events (cf. Ricoeur 1990, third section, pp. 73, esp., for example, p. 93).



and in this way one angel speaks to another; for to speak to another only means to make known the mental concept to another. (S.th. I, q. 107, a. 1, c.)<sup>13</sup>

However, one could ask in a critical vein whether this description is really meaningful. What could the contents of angelic communications be if their knowledge is not grounded in a perspective of situated bodily existence? It is not convincing that angels have to let other angels know something because angels already possess an identical knowledge. A better explanation might point out that Aquinas ascribes free will to the angels, so that angels are distinguishable in their will. This is the reason for a kind of privacy of their thoughts that cannot be immediately known by other angels but may be communicated to them (cf. Goris 2012, p. 182; cf. Roling 2012, p. 236). However, from a phenomenological standpoint, it is hard to clarify how an angelic communication based on their different wills is possible. In any case, this communication could not be compared in any way to human communication which is interested in a social exchange and in a social participation that transcends our personal and individual limitations. For in these cases we would have to introduce a situated will of an angel which stands in contrast to the aforementioned considerations.<sup>14</sup>

If we use these thoughts about angels to reflect on human agency, even social context turns out to be a constitutive condition of action. To illustrate this I will take a look at normal situations in which we talk about actions, like – for example – a quarrel about the question of whether another one hit me intentionally or only accidentally (cf. Mertens 2013, p. 148). This means: we want to clarify whether someone has acted in the proper sense of the term or whether he or she merely behaved. Such situations demonstrate the following: Whenever an action is performed and we talk about it, we presuppose that *something* was or is done. Even though it might sometimes be difficult to tell whether the thing done was an action at all and even difficult to identify exactly *what* was done, as soon as we call something an action, we identify it in specific regards or at least try to understand it as an action which is defined in certain ways. Actions are specified as actions. This demonstrates that, in a broad sense, actions have normative character, which means that actions are defined by certain rules an acting person has to comply with. And, as a matter of principle, we never decide about such

---

**13** “Ex hoc vero quod conceptus mentis angelicae ordinatur ad manifestandum alteri, per voluntatem ipsius angeli, conceptus mentis unius angeli innotescit alteri: et sic loquitur unus angelus alteri. Nihil es enim aliud loqui ad alterum, quam conceptum mentis alteri manifestare.”

**14** It should be mentioned that researchers of Aquinas probably would not share the former critique. Goris, for instance, states that “there is not a fundamental, only a gradual, difference between humans and angels” (Goris 2012, p. 183).

rules on our own – as we can say following Ludwig Wittgenstein and Peter Winch. Following rules is a public matter. By communicating with others we define what it means to perform this and not that action; we ascribe actions to one another clarifying why person A and not person B is responsible for a specific action. And we also find categories which can be applied to judge the specific way of performing an action. In short: communication about our behaviour is an essential precondition of our concept of action. As an attempt to transcend the perspectives of the participants, communication presupposes the bodily situated existence of the involved individuals.

I want to finish by pointing out the main conclusion of my considerations: Both the social and individual sources of our concept of action presuppose corporeally-situated agents like us. Therefore, it is no coincidence that the talk of angelic actions seems to be fundamentally flawed – at least to our human ears. We simply cannot imagine what it is like to be an angel. However, by dealing with the concept of incorporeal entities like angels we may gain relevant insights in fundamental structures of our own experience, particularly as acting persons.<sup>15</sup>

## Bibliography

- Goris, Harm (2012): “Angelic Knowledge in Aquinas and Bonaventure”. In: Tobias Hoffman (Ed.): *A Companion to Angels in Medieval Philosophy* (Brill’s Companions to the Christian Tradition, vol. 35). Leiden, Boston: Brill, pp. 149–185.
- Leist, Anton (2007): “Introduction: Through Contexts to Actions”. In: Anton Leist (Ed.): *Action in Context*. Berlin, New York: Walter de Gruyter, pp. 1–51.
- Mertens, Karl (2013): “Soziale Dimensionen der Normativität. Perspektiven einer phänomenologischen Analyse handlungskonstitutiver und sozialer Normen”. In: Dieter Lohmar / Dirk Fonfara (Eds.): *Soziale Erfahrung* (Phänomenologische Forschungen 2013). Hamburg: Felix Meiner Verlag, pp. 145–164.
- Nagel, Thomas (1974): “What Is It Like to Be a Bat?”. In: *The Philosophical Review* 8, pp. 435–450; again in: Nagel, Thomas: *Mortal Questions*. Cambridge: Cambridge University Press, first published 1979, 11<sup>th</sup> printing 2006, pp. 165–180.
- Ricœur, Paul (1990): *Soi-même comme un autre*. Paris: Édition du Seuil.
- Roling, Bernd (2012): “Angelic Language and Communication”. In: Tobias Hoffman (Ed.): *A Companion to Angels in Medieval Philosophy* (Brill’s Companions to the Christian Tradition, vol. 35). Leiden, Boston: Brill, pp. 223–260.

---

<sup>15</sup> I am very indebted to Therese and David Cory as well as Jörn Müller who corrected my English text and gave me valuable hints regarding the topic of Aquinas’s angelology. In addition I am much obliged to Michela Summa for linguistic and philosophical critique of an earlier version of this contribution.

Strawson, Peter Frederick (2006), *Individuals. An Essay in Descriptive Metaphysics* (1<sup>st</sup> ed. 1959). London, New York: Routledge.

Thomas Aquinas (S.th) (1988): *Sancti Thomae de Aquino Summa Theologiae*, Milan: Edizioni Paoline.

Thomas Aquinas (S.th.) (no year): *Summa Theologica* (Christian Classics Ethereal Library) <http://www.ccel.org/ccel/aquinas/summa.pdf> [Last accessed 12/2/2015].



## Social Philosophy and Collective Intentionality



Thomas Szanto

# Do Group Persons have Emotions – or Should They?

**Abstract:** Group person, if there are any, are not made of flesh and blood. Can they, nonetheless, have an emotional life of their own? But why should they in the first place? And what if they cannot? Are, then, group persons ‘fictional’ entities or some robotic ‘artificial’ person? Finally, do corporate entities that lack certain affective abilities necessary for moral conduct, moral address or accountability, such as fellow-feelings or sympathy, exhibit psychopathic behaviour, as some have indeed argued? These are the questions I wish to address in this article. I will do so by, first, outlining the main positions in the contemporary socio-ontological and moral-philosophical discussion on group personhood, and showing how, surprisingly, almost none of the available accounts pay any attention to the question of whether emotions, of some sort, may be attributed to group persons (section 2). I will, then, explore four central reasons that could motivate to pursue the task of attributing emotions to GP in the first place (section 3). Next, I will reply to the two most apparent reasons that could militate against such an attribution (section 4). I shall conclude that there seems to be no principled reason to repudiate the possibility of group persons who may in fact exhibit some form of sensibility—whatever emotional reasons there may well be for being afraid of them.

## 1 Introduction

Group persons, if there are any, are not made of flesh and blood. Can they, nonetheless, have an emotional life of their own? But why should they in the first place? And what if they cannot? Are, then, group persons only “soulless”, “fictional” entities, as the Scholastics already declared (Kantorowicz 1957), or some robotic, or “artificial” person, as Hobbes would later famously put it? Finally, do corporate entities, which lack certain affective abilities necessary for moral conduct, moral address or accountability, such as fellow feelings or sympathy, exhibit psychopathic behaviour (Bakan 2004)? These are the questions I wish to address in this article. I should note here that, although I do not intend the title of this paper to be taken simply as a rhetorical question, I will not offer any conclusive answer to it either. What I wish to do, instead, is to think more seriously about the conditional: what kind of ontological and normative concep-

tion of group personhood would result if we either accepted or rejected the view that group persons have emotions.

I will first do so by sketching the contemporary socio-ontological and moral-philosophical discussion on group personhood and showing how, surprisingly, almost none of the available accounts pay any attention to the question of whether emotions, of some sort, may be attributed to group persons (section 2). I will then explore the most important reasons that could motivate such an attribution (section 3), and conclude by replying to the two most apparent reasons that could militate against such an attribution (section 4).

To get an initial flavor for the issue at stake, consider the following type of statements that we can regularly encounter in the media: ‘G.M. is concerned about the steep rise of oil prices’; ‘Germany deeply regrets the war crimes committed in Greece.’ Such statements have also been recently analyzed by a number of authors working on collective emotions (e.g., Gilbert 2002; Tollefsen 2003; Huebner et al. 2010; Huebner 2011; Gilbert 2014; Schmid 2014). No matter how one specifies their subject (as legal, artificial, etc.), such statements are not only attributed by third parties but are often also self-attributed, as in the following self-promotion issued by a computer company: “HP is a firm where one can breathe the spirit of interrelations [...]. It is an affective relationship” (cited from Illouz 2007, p. 22).

Now, it seems little contentious that the mental or psychological states attributed in such sentences or self-ascribed in the above statements are emotions or affective states (regret or worrying), or are laden with some emotional import (being concerned about). Moreover, many today would agree that these and many similar statements have group or corporate persons as their subjects. Let’s, then, take a closer look at the current discussion.

## 2 Mapping the contemporary landscape

In the past few decades, a number of philosophers working on the metaphysics of personal identity have raised the question whether individuals can integrate into a single personal unit that differs from ordinary, human-sized persons (Parfit 1986, p. 211; Korsgaard 1989; Rovane 1998). The issue of group personhood has also been considerably fueled by debates on collective intentionality and group agency (see references below).<sup>1</sup> Further, in times of ever more concentrated corporate power, one can witness a growing interest in the respective norma-

---

<sup>1</sup> See Szanto 2016a and 2016b for a phenomenological reassessment of the current discussion.

tive and political questions. For example, some have inquired whether we have proper obligations towards group persons, or whether they have, above and beyond moral accountability (French 1979; Manning 1984), any moral or political rights of their own, such as the right for protection, the right to persist, or the right to freedom of speech (Ozar 1985; Stoll 2005; List & Pettit 2011; Briggs 2012; Hess 2013; Hindriks 2014; Kusch 2014).<sup>2</sup>

Unsurprisingly, there is no consensus regarding what properties, rights or obligations entities must bear in order to count as persons, or what the necessary and sufficient conditions of individuating persons are. A number of proposals have been made, which often enough either confound or else deliberately mix cognitive or rational properties and practical, deliberative, or agential, as well as moral, faculties. The situation is not much different when it comes to *group* persons (henceforth GP). Beyond merely legal, business ethics, and sociological or political theory conceptions of corporate personhood,<sup>3</sup> in the barest outline, we may discern the following four main types of socio-ontological GP-construals in the current literature:<sup>4</sup>

1. *The Moral Address and Moral Agency Attribution Account*: Though ontologically there are no proper GP (or, alternatively, whether or not there are such), a collective C can, and in fact should, be counted and treated as a GP, if C is the appropriate target of moral address. Thus, if (i.) C's morally relevant joint action is irreducible to a subset of or the aggregation of the individual members' actions, and if (ii.) either the moral practice of attributing (collective) responsibility to C, or the interpersonal practice of entertaining so-called reactive attitudes (à la Strawson 1968), such as resentment, gratitude, forgiveness, blame, etc., towards C, are justified, then C should be treated as a (characteristically personal) subject bearing moral agency, accountability and responsibility (Finnis 1989; Tollfesen 2003; Graham 2004; see also with slightly different emphasis, Manning 1984; Sheehy 2006; Huebner 2014; cf. Hess 2013).

---

<sup>2</sup> This is particularly pertinent after the 2010 *Citizens United v. Federal Election Commission* US Supreme Court ruling, according to which corporations are entitled to free speech rights under the First Amendment; see: <http://www.supremecourt.gov/opinions/09pdf/08-205.pdf>, cf. Hess 2013.

<sup>3</sup> Think of the most influential types such as the “realist”, the “fiction” or the “legal aggregate” theory of corporate entities; cf. French 1979; Hager 1989; Bratton 1989; Runciman 1997. From the veritable industry of historical systematic legal- and political-scientist treatments on this issue, see Dan-Cohen 1986; Mark 1987; Hirst 1989; Blumberg 1993; Ireland 1996, and Wendt 2004.

<sup>4</sup> Consider that the issue whether persons, *qua* social beings, are part of our social ontology is, though obviously related, orthogonal to the issue of whether that ontology includes group persons; for a positive recent argument of the former issue, see Baker 2015.



2. *The Moral Personhood Account*: Certain groups and corporations are, ontologically, proper persons, if they are intentional and moral agents who are licensed by having a “Corporation’s Internal Decision Structure” (CID). CID is a “responsibility flow chart” that, together with certain corporate decision rules embedded in corporate policies and recognized as such, determines the levels of, and ratification, exercise, and control of agency within, a corporation’s power structure (French 1979).
3. *The Collective Self-Concept Account*: A collective has “person-like properties”, such as the capacity to intend, act, plan, regret, or hope, if its members take a first person plural perspective and have a “shared self-concept” (which also includes membership-criteria and is common knowledge among members) (Mathiesen 2003).
4. *The Unity of Rational Agency and Social Integrate Account*: A collective qualifies as a group person if it is a unified intentional agent that has its own “rational point of view” (Rovane 1989), from which it deliberates, and which guides its actions and the formation of new beliefs and commitments, and that may well be discontinuous with the rational or “deliberative standpoint” (Korsgaard 1989) of each and every member of the group. Moreover, such “social integrates” (Pettit 2003) can be collectively held (morally) responsible for the failure to rationally unify or integrate their actions and beliefs (see also List & Pettit 2011; Szanto 2014).

Clearly, these are not incompatible accounts. On the contrary, as should strike the reader obvious, there are various interrelations between them, and especially between (2), (3) and (4). One may also lay the emphasis on some other typical properties of personhood that distinguish corporate persons, such as voluntariness of association, remorsefulness for moral transgression, openness and answerability (Scruton 1989), or “conversableness” (Pettit 2003; Pettit/List 2011). Notice also that only (1) represents a sort of (Dennettian or Davidsonian) “interpretationism” that, despite taking the attribution of (moral) personhood to collectives to metaphysically be an open question, or to be unjustified even, still sees it as necessary for normative and explanatory reasons (cf. Tollefsen 2003). All other accounts grant collectives the proper metaphysical status of being persons. There are also some possible intermediate positions between (1) and (2): e.g., one may hold that GP are, though moral agents, no full-fledged persons, because they have no full-fledged moral (but only conventional, or collectively accepted) rights (Ozar 1985). Alternatively, one may hold that the important question is not metaphysical but, rather, whether groups ought to be treated as individual moral persons, and whether we should have the same scruples about

being fair to them or holding them morally responsible as we do in the individual case (Manning 1984; see more below, sec. 4).

Meanwhile, there is very little explicit opposition to the attribution of personhood to collectives – which is curious enough, given that, even among rather hard-boiled realists about group agency, there is only a handful who are, however scarcely, willing to make such attributions explicit.<sup>5</sup> Those who explicitly argue against GP typically deny (1) to (4) altogether, doing so either precisely on the ground that collectives are not moral agents (Ranken 1987), and hence not persons or, conversely, that they are not persons, and hence are proper bearers of neither moral rights, nor agency (cf. Wellman 1999). Finally, we have those who—by explicitly distinguishing between a bearer of moral rights, moral agency, and personhood—either claim that, though corporate entities are not persons, and hence have no moral rights proper, they do ‘inherit’ all necessary properties of moral agency from the respective individuals (Neuhäuser 2011), or that they are full-fledged rational, intentional and moral agents, but still not proper persons, because personhood entails some further properties that collectives lack, notably emotions, affectivity, or the capacity to experience vulnerability (Hess 2013; cf. Manning 1984).

In any case, notice the almost complete lack of emotional attribution to GP. Thus, I am not aware of a single author who would explicitly attribute emotions or affective sensitivity to GP, including, notably, those social ontologists who indeed argue for shared, collective or corporate emotions (e. g., Salmela 2012; Gilbert 2014; Schmid 2014). Indeed, I only know of a single author who has directly argued against GP’s being persons precisely on the grounds that they do not exhibit any emotional vulnerability or sensitivity (Hess 2013). (I come back to that in sec. 5 below.) Conversely, I know of only one prominent argument *in favor of* GP, but it still conceives of GP as lacking sensitivity, in any personhood-defining sense (Rovane 1998). Incidentally, this almost complete ignorance fairly reflects a much broader tendency in the philosophy of personhood: the question whether, and in what sense, personhood requires emotions or affectivity has hitherto been paid quite generally very little attention.<sup>6</sup>

---

<sup>5</sup> For a typical example of a realist about group agency and anti-realist about group personhood, see Tuomela’s “non-entity view” of group agents: “Groups are not persons, because they have neither bodies nor minds” (Tuomela 2007, 145). Notice that this is *all* Tuomela has to say on group personhood.

<sup>6</sup> Notice that, in this paper, I cannot dwell upon any rational, normative, or any other non-emotional, general requirements that an adequate theory of GP must fulfill; I have done so in Szanto 2015, where I elaborate on what I label “plurality”, the “integrity”, “normativity, or moral accountability” and “anti-collectivism” requirement respectively. Against this backdrop, the pres-

### 3 Why get emotional about group persons?

At this point, skeptics, however, might raise the following general concern: why should one pursue the task of attributing emotions to GP, or even just raise this issue, in the first place? After all, both the issue of group personhood and that of collective emotions are contentious enough when discussed separately. Linking these two issues will, so the worry might go, make things even more contentious rather than resolve anything. So why bother about whether groups or corporate entities have emotions? I contend that there are a number of reasons for raising the issue, independently of whether we ultimately attribute emotions to GP.

(1) First, consider how our conception of what it is to be a person is deeply entangled with our conception of what it is to be a subject that has an emotional life. This (intuitive) conceptual entanglement obtains even if, as mentioned, philosophers, rather curiously, remain almost completely silent on spelling out exactly this connection.<sup>7</sup> This is especially apparent when it comes to moral psychology – and I shall come back to that in a moment – but I think that the conceptual entailment between emotional sensitivity and personhood is not limited to moral emotions, such as shame, forgiveness, indignation, or pride (Steinbock 2014, 274–275). But our web of emotions not only defines human subjectivity and, in large part, an individual's *personality*, or her moral or non-moral character (Rorty 1978; Solomon 1993, p. 70 f.; Ben-Ze'ev 2000, pp. 150–157; Goldie 2000, pp. 141–175). It is our affective dispositions and abilities, their assessment and their education, or our very sensitivity and emotional vulnerability that also lie at the heart of what it means to be a person. Moreover, it is shared focuses of import that define many human communities. Without any emotional import, concern or care, not only would there be no romantic couples, no fan clubs, religious communities, reading clubs, but most natural communities, like families, would also lose much of their cohesive power. (To be sure, not all such communities are, or act as, GP, but some, arguably, do, e.g., a fan club that “personifies” itself in choosing a spokesperson). As Goldie writes about the connection between emotions and what he aptly calls the “personal point of view”, which is opposed not to the objective, third-personal but, rather, to an “impersonal” one: “our ordinary everyday thought and talk about the emotions and emotional experiences is essentially from the personal point of view”

---

ent paper is an attempt to get clearer about whether we need a further requirement concerning some emotional capacities (see next sec.).

<sup>7</sup> See, however, authors in the phenomenological tradition, and above all Scheler's work (e.g., 1926), but also that of Amelie Rorty (e.g., 1978).

(Goldie 2000, 3; cf. p. 82f., pp. 181ff.). Surely, this – however essential or constitutive – connection between emotions and persons does not *per se* license the inference that all types of persons necessarily have emotions; having said this, however, our options seem, intuitively, limited to either of the following alternatives: (a) if we are reluctant to consider seriously the option of attributing some collective or corporate emotions to GP, we should a.) either stop talking of group persons (in which case, i.e., if we are reluctant to posit any GP in our social ontology, it may be better not to employ the talk of collective emotions at all); (b) or we should reassess our notions of (reactive) emotion and personhood altogether, and devise a theory according to which these two personhood and emotional properties may fundamentally come apart, as, so the suggestion might go, they do in the collective case.

(2) Now, whatever psychological or metaphysical theory of personhood or GP we employ, and however emotions factor in there – I believe the single most important reason to attribute emotions to GP is due to the moral address account and, in particular, to the role reactive attitudes play therein (cf. also Tollefsen 2003; Helm 2008; 2014). Incidentally, a large number of authors today consider reactive attitudes to be definitive of moral agency and responsibility in the individual case, too (cf., e.g., Vargas 2013).

First, it may be argued that reactive attitudes need emotions to reliably track them, or to have some morally motivational power, even if this turns out to be a contingent fact of human (moral) psychology. Mele has put this point nicely:

Perhaps, owing to our basic psychological make up, we simply cannot learn to regard the interests of others as at all important unless we are sensitive to emotions of other people [...] and learn to see ourselves as apt targets of some reactive emotions. Perhaps such sensitivity and learning are, for us, psychologically necessary for the acquisition of any sort of moral sensitivity or understanding. This is a contingent matter. [...] Might a hypothetical, isolated community of emotionless beings –hence, beings with no reactive attitudes – include morally responsible agents? (Mele 2000, p. 449)

The idea here is that, if you weren't able to feel ashamed when being blamed, you wouldn't be able to be subjected to reactive attitudes such as blame either, nor would you be able to see when it is appropriate to blame somebody else; moreover, we wouldn't know when and why it would hurt somebody else to be blamed (unjustly or not). It is also this emotional-*cum*-moral sensibility, and indeed sensitivity, that we both draw upon and try to modulate when we educate the interpersonal behaviour of our children. Now, there seems to be no principled reason why such moral-*cum*-emotional sensibility should not be possible and, indeed, appropriate in the collective, or group-personal, case.

Secondly, one might conjecture that such a sensibility would allow to faster or more efficiently track the appropriateness of certain actions vis-à-vis certain corporate values than would highly complex corporate decision making procedures (cf., Schmid 2014). In fact, in the individual case, it has been argued that emotions facilitate more intuitive, effortless, and more creative cognitive processes, though, unsurprisingly, often at the detriment of cognitive control and vigilance (Kahneman 2011, esp. pp. 67–70, 137–145).

But let's set aside for a moment these points about moral motivation, and our contingent cognitive psychology. There is, thirdly, another systematic argument for a structural link between emotions and reactive feelings, or attitudes, that remains. It is related to the moral-psychological reasons mentioned above, and has to do with the *interactive* nature of reactive feelings, precisely *qua* reactive. Thus, we cannot have any reactive feelings or attitudes in a one-way manner. They necessarily require that both parties involved are creatures that are able to form such attitudes and understand what they imply. Moreover, even if they only represent or target *moral* emotions (guilt, resentment, blame, etc.), and even if they are not necessarily accompanied by bodily changes or bodily feelings<sup>8</sup>, reactive attitudes certainly do require emotional attitudes on the part of the subject whom these attitudes are directed at. For example, blame would be an utterly inadequate reactive emotion towards stones, but it is also inadequate if the person at whom it is directed does not herself *feel* accordingly in some way or another. For instance, we expect that somebody will feel ashamed or guilty when being blamed, proud, or even embarrassed, but not afraid, when being praised, etc. In other words, what is partly, but essentially so, targeted when we entertain reactive attitudes towards somebody is that the person feels *what* she does, and *how* she does, according to certain moral, but also affective, norms and standards.

(3) Relatedly, to entertain reactive attitudes arguably also requires creatures who have the capacity to exercise empathy and, eventually, sympathy with others, and this capacity, in turn, will require some sort of emotional or affective life. Consider that this claim does not depend on a more specific doctrine according to which empathy is not primarily a cognitive but an affective state, and/or that it requires some affective interpersonal similarity between empathizer and target (Jacob 2011). Moreover, neither does this claim depend on any claims concerning empathy amounting to or presupposing some affective sharing (e.g., Hein & Singer 2008). Following a number of phenomenologists (e.g., Zahavi 2011, p. 2014), I do not endorse either of these claims (cf. Szanto 2015). The point I

---

<sup>8</sup> See more on this issue below, sec. 4.

wish to make here, however, is more general: arguably, only creatures with *some sort or another* of an emotional or affective life will be able to empathically experience other emotional creatures, or empathically access and understand what it is to have (any) emotions in the first place; this is also even the case when one empathically understands or sympathizes with specific emotions that one has never experienced (e. g., the grief for the death of another’s parents’ while one’s own are still alive).

At this juncture, one might, of course, object that empathy is not necessary for morality. For example, Prinz (2011) has recently challenged the view that empathy is or even *should* be viewed as necessary for making moral judgments, for moral development, or even for motivating moral conduct. Consequently, he argues that, in education, socialization, etc., we *should* not even cultivate any empathy-based morality. Now, Prinz defines empathy, all-too narrowly and ultimately erroneously, in my view, as “a kind of vicarious emotion: it’s a feeling what one takes another person to be feeling” (Prinz 2011, p. 212). But this is not the main issue here. Neither is the moot point that we need empathy in order to share some emotions and eventually sympathize with them. Rather, empathy, understood as the ability to access others’ mental and affective states, will be necessary if one wishes to retain the reactive attitude account of (moral) agency. For consider again that reactive attitudes can only have any moral motivation and significance for agents who have the capacity to empathically access, understand, and assess reactive attitudes and feelings (e. g., knowing what it is to forgive, to be blamed, or to feel guilty). But, to repeat, only creatures with some emotional abilities will be endowed with such rich forms of interpersonal understanding.

(4) Lastly – and probably least controversially, but also psychologically and metaphysically least interestingly – in the past couple of decades, one can witness in organizational psychology and management an ever deeper, and ‘muddier’, entanglement between ‘corporate cultures’, or the set of norms and values governing companies and institutions, and what is sometimes called ‘emotional cultures’. In her intriguing sociological work, Illouz has elaborated this entanglement in terms of what she labels ‘emotional capitalism’ (Illouz 2007). Consider also in this connection a number of organizational studies investigating the correlation between certain negative corporate emotional cultures, such as bullying, unfair supervision or the extreme monitoring of employees, on the one hand, and the presence of so-called “corporate psychopaths” or “collective paranoia” (e. g., Kramer 1994; Boddy 2011), on the other. Notice, further, that a corporate emotional culture, if there is any, is not the same as what is sometimes used synonymously with genuinely collective, or group-level, emotions and labeled “corporate emotions” (e. g., Schmid 2014)

At this point, one will surely want to know what kind of emotions we are actually talking about: what kind of emotions, precisely, are those that might – if at all – be attributed to GP? I can neither dwell upon the exact nature or constitution of collective emotions or the norms governing them, nor on how they might differ either from some less robust forms of emotional sharing, or else from emotional contagion (see Szanto 2015). Instead, against the backdrop of the, admittedly sketchy, considerations so far, I want to suggest the following requirement, call it the *Corporate Emotional and Moral Sensibility Requirement*, for there being, potentially, any GP-type emotions:<sup>9</sup>

(CEMS) A group person GP may, but will not necessarily, have GP-type emotions if it has a ‘shared emotional culture’ with a robust evaluative and normative, or ‘social appraisal’, structure, *within* or *relative to* the GP, and if its members have direct perception-based or other mediated mutual awareness of there being such a shared emotional culture. Moreover, members care, or ought to care, for the maintenance of this culture and for the appropriateness of the social appraisal structure vis-à-vis the shared or corporate moral and non-moral values and stances.

## 4 Objections and replies

Let me close this paper by replying to the two most apparent lines of objection that critics might raise – not so much against this requirement (CEMS) itself but, rather, against any properly group-personal emotions instantiated, *provided* that CEMS is fulfilled.

The first line of objection to any account of corporate emotions seems to be that GP are not in any straightforward sense embodied agents, and hence cannot be said to have emotions, understood as bodily changes or feelings. There is not much to say to ward off this objection—as long as we take emotions to be reducible to bodily feelings. But such a reduction is surely not mandatory. For one, not only most social psychologists (e. g., Parkinson et al. 2005), but also most philosophers today, and not only emotional cognitivists (e.g., Solomon 1993; Nussbaum 2001), or even those who indeed stress the importance of bodily changes or embodied emotional behaviour (e.g., Damasio 1994; Prinz 2004; Colombetti 2014), would argue for a much richer notion of emotions than one merely involv-

---

<sup>9</sup> Again, consider that there are further necessary requirements for there being any GP in the first place (see above, FN 7, and Szanto 2016b). For a more elaborate exposition of the respective natures of the related, but different, shared and collective emotions, see Szanto 2015, and of social appraisal, in particular, Bruder et al. 2014.



ing bodily feelings (e. g., Ben-Ze'ev 2000; Goldie 2000; Roberts 2003; Deonna & Teroni 2012). Indeed, and more to the point, when it comes to the emotional domain of group persons, the emotions at stake will precisely not be such embodied affective states as, for instance, trembling from fear, but more complex, typically moral, emotions, such as being socially humiliated, not being recognized as a group, or some collective guilt feelings concerning harms committed by one's own group, and the like.

Relatedly, critiques might object that corporate persons cannot, properly speaking, *suffer*. For example, one might argue that, if a GP is actively 'terminated', or otherwise ceases to exist or, if it is, as it were, put in some painful, or even just some awkward or embarrassing, position, in neither of these cases, if conceivable at all, will GP suffer. Whether or not one, eventually, deduces from this difference between GP and individual persons any differences in their respective rights—as one typically does for persons and animals versus plants or robots—for some, this difference regarding the vulnerability of human versus corporate persons will seem crucial. Obviously, however, the issue depends on what we take suffering to entail or presuppose. On one reading, suffering is simply associated with some bodily feelings (e. g., physical pain). It is in this sense that some claim that it makes no sense to expand, for instance, the prohibition of (physical and psychological) torture to corporations and group agents (Briggs 2012). We find the most sustained elaboration of this objection in a paper of Hess (2013), in which she argues that corporate entities are, though potentially intentional-*cum*-rational and moral agents, not persons, and in which she appeals to the distinctive (emotional) vulnerability of persons to make her case:

It is our vulnerabilities, as much or more than our powers, that are both distinctive and constitutive of our personhood. It is only because we persons have the specific vulnerabilities that we do—to physical and emotional suffering, to worry and weariness, to humiliation and despair—that we are entitled to or even in need of the specific protections typically granted to us. There is thus no justification for assigning those protections to corporations or anything else that lacks them. (Hess 2013, p. 335)

But, fortunately, there is another reading of vulnerability available than that regarding sheer physical or bodily-affective vulnerability: for it might, just as well, entail moral or some other more complex kind of suffering (apart from, trivially, and particularly so for corporations, financial losses or material suffering). True, group persons will, for example, not 'despair' in the exact same ways as individuals; but communities, and by extension, arguably, GP, might very well, for example, be humiliated (cf. Honneth 1994; Copp 2002). Consider a minority that would very well 'despair', namely by having recognition withheld to the point of resorting to terrorism, or one that is humiliated, for instance, by means of



one of its members' being publicly ridiculed for her group membership. Or consider an employee who is humiliated *qua* employee of a given company, or a family, or other strongly cohesive community, who genuinely *fears* being extinguished or dying out, where this is felt collectively, and *for* the collective, as a family, or as a minority.

(2) The second line of objection is closely related to the first. According to this concern, emotions and feelings, just like suffering, incidentally, require that the subject who has them has some phenomenal consciousness, or is consciously aware of having them. After all, emotions, and especially their often-invoked 'qualitative feel', seem, in a more straightforward sense than any other type of mental state,<sup>10</sup> to be paradigmatic examples of conscious states. Moreover, one may, and rightly so, question whether group persons, or group *minds*, must be collectively conscious or not; indeed, I have argued elsewhere that they need not and, indeed, must not (Szanto 2014). In this vein, for example, Huebner (2011) has voiced similar skepticism. Instead of phenomenally conscious (group-)personal level collective emotions, he argues, rather, for *subpersonal* level collective emotional processes, which may serve similar cognitive functions as individual (and, possibly, personal) level emotions.

But, again, we have a strategy available that allows for reactive emotions to be dissociated from such instances of phenomenal consciousness. According to this, reactive emotions are not individuated by their phenomenology – which GP may then well lack—but, rather, by the norms governing them. Here is how Tollefsen puts this point:

The fact that collectives lack phenomenology, then, does not mean that they lack the capacity for reactive attitudes. [...] Collective reactive attitudes would be differentiated on this approach from individual reactive attitudes in terms of the norms governing the relevant reactions. Certain emotional responses might be licensed for an individual only because of her group membership. They would also be differentiated in terms of their motivational upshot. The 'pangs' of remorse you feel *qua* group member may lead you to do actions that you would not do *qua* individual. Although this approach does not have the collective itself feeling the emotion, it does identify a way in which emotions can be collective. Perhaps this is all we need in order for collectives to be appropriate targets of our reactive attitudes. (Tollefsen 2003, p. 232)

---

**10** This is no way to deny or downplay cognitive phenomenology, or the essential correlation, stressed especially by phenomenologists such as Husserl or Sartre, between pre-reflective self-awareness, phenomenal consciousness and the intentionality of mental states; see Szanto 2012.

## 5 Conclusion

Following this rather cautious move here, and in concluding, let me restate the initial restriction: the argument of the paper should only be understood as a conditional, exploring what would be – descriptively, and, in particular, normatively – gained or lost if we were to attribute a certain sensibility or corporate emotions to group persons. Moreover, let there be no misunderstanding: just as we ought to be as parsimonious as we can in attributing personhood to collectives, so ought we to be very considerate when attributing emotions to group persons. In any case, many types of group agents and collectives, and especially corporations, will be better explained without attributing any emotions to them. They may still pursue goals, or have a rational point of view from which they deliberate or take decisions, even without having a more integrated *phenomenal* point of view, or focus of import and care. Similarly, we should be more careful in distinguishing different conceptions of personhood when it comes to collectives, just as we should be careful not to see GP where we only have collective emotions, or where we even only have some corporate emotional cultures without any GP. On the other hand, there seems to be little philosophical, or incidentally, normative, reason to be afraid of group persons who may in fact exhibit some form of sensibility—whatever emotional reasons there may well be for being afraid of them.

## Bibliography

- Bakan, Joel (2004): *The Corporation. The Pathological Pursuit of Profit and Power*. New York: Free Press.
- Baker, Lynne R. (2015): “Human Persons as Social Entities”. *Journal of Social Ontology* 1(1), pp. 77–87.
- Ben-Ze’ev, Aaron (2000): *The Subtlety of Emotions*. Cambridge, MA: MIT Press
- Blumberg, Philip I. (1993): *The Multinational Challenge to Corporation Law. The Search for a New Corporate Personality*. Oxford: Oxford University Press.
- Boddy, Clive R. (2011): “Corporate Psychopaths, Bullying and Unfair Supervision in the Workplace”. *Journal of Business Ethics* 100(3), pp. 367–379.
- Bratton, William W. (1989). The New Economic Theory of the Firm: Critical Perspectives from History. *Stanford Law Review*, 41(6), pp. 1471–1527.
- Briggs, Rachael (2012): “The Normative Standing of Group Agents”. *Episteme* 9(3), pp. 283–291.
- Bruder, Martin/Fischer, Agneta/Manstead, Antony S. R. (2014): “Social Appraisal as a Cause of Collective Emotions.” In: C. von Scheve/M. Salmela (Eds.): *Collective Emotions*, Oxford: Oxford University Press, pp. 141–155.
- Colombetti, Giovanna (2014): *The Feeling Body. Affective Science Meets the Enactive Mind*. Cambridge, MA: MIT Press.

- Copp, David (2002): "Social Unity and the Identity of Persons". In: *The Journal of Political Philosophy* 10(4), pp. 365–391.
- Damasio, Antonio (1994): *Descartes' Error. Emotion, Reason, and the Human Brain*. London: Penguin.
- Dan-Cohen, Meir (1986): *Rights, Persons, and Organizations: A Legal Theory for Bureaucratic Society*. Berkeley: University of California Press.
- Deonna, Julien A., & Teroni, Fabrice (2012): *The Emotions. A Philosophical Introduction*. London/New York: Routledge.
- French, Peter A. (1979): "The Corporation as a Moral Person". In: *American Philosophical Quarterly* 16 (3), pp. 207–215.
- Finnis, John (1989): "Corporate Persons". In: *Proceedings of the Aristotelian Society, Supplementary Vol.* 63 (1989), pp. 266–274
- Gilbert, Margaret (2002): "Collective Guilt and Collective Guilt Feelings". In: *The Journal of Ethics*, 6(2), pp. 115–143.
- Gilbert, Margaret (2014): "How We Feel: Understanding Everyday Collective Emotion Ascription." In: C. von Scheve/M. Salmela (Eds.): *Collective emotions*. Oxford: Oxford University Press, pp. 17–31.
- Goldie, Peter (2000): *The Emotions. A Philosophical Exploration*. Oxford: Oxford University Press.
- Hager, Mark M. (1989): "Bodies Politic: The Progressive History of Organizational 'Real Entity' Theory". In: *University of Pittsburgh Law Review* 50(3), pp. 575–654.
- Helm, Bennett (2008): "Plural Agents". In: *Noûs* 42(1), pp. 17–49.
- Helm, Bennett (2014): "Emotional Communities of Respect." In: C. von Scheve, & M. Salmela (Eds.): *Collective Emotions*. Oxford: Oxford University Press, pp. 47–60.
- Hess, Kendy M. (2013): "'If You Tickle Us...': How Corporations Can Be Moral Agents Without Being Persons." *The Journal of Value Inquiry* 47(3), pp. 319–335.
- Hindriks, Frank (2014): "How Autonomous Are Collective Agents? Corporate Rights and Normative Individualism." *Erkenntnis* 79(9), pp. 1565–1585.
- Hirst, Paul Q. (1989): "Introduction." In: P. Q. Hirst (Ed.): *The Pluralist Theory of the State. Selected Writings of G.D.H. Cole, J.N. Figgis, and H.J. Laski*. London/New York: Routledge, pp. 1–45.
- Honneth, Axel (1994): *Kampf um Anerkennung: Zur moralischen Grammatik sozialer Konflikte*. Frankfurt a.M.: Suhrkamp.
- Huebner, Bryce (2011): "Genuinely Collective Emotions". In: *European Journal for Philosophy of Science* 1(1), pp. 89–118.
- Huebner, Bryce (2014): *Macrocognition. A Theory of Distributed Minds and Collective Intentionality*. Oxford: Oxford University Press.
- Huebner, Bryce/Bruno, Michael/Sarkissian, Hagop (2010): "What Does the Nation of China Think About Phenomenal States?". In: *Review of philosophy and psychology* 1(2), pp. 225–243.
- Illouz, Eva (2007): *Cold Intimacies. The Making of Emotional Capitalism*. Cambridge: Polity.
- Ireland, Paddy (1996): "Capitalism Without the Capitalist: The Joint Stock Company Share and the Emergence of the Modern Doctrine of Separate Corporate Personality". In: *The Journal of Legal History* 17(1), pp. 41–73.
- Kahneman, Daniel (2011): *Thinking, Fast and Slow*. London: Penguin.

- Kantorowicz, Ernst H. (1957): *The King's Two Bodies: A Study in Medieval Political Theology*. Princeton: Princeton University Press 1997.
- Korsgaard, Christine M. (1989): "Personal Identity and the Unity of Agency: A Kantian Response to Parfit". In: *Philosophy & Public Affairs* 18(2), pp. 101–132.
- Kusch, Martin (2014): "The Metaphysics and Politics of Corporate Personhood". In: *Erkenntnis* 79(9), pp. 1587–1600.
- Kramer, Roderick M. (1994): "The Sinister Attribution Error. Paranoid Cognition and Collective Distrust in Organizations". In: *Motivation and Emotion* 18(2), pp. 199–230.
- List, Christian/Pettit, Philip (2011): *Group Agency. The Possibility, Design, and Status of Corporate Agents*. Oxford: Oxford University Press.
- Mathiesen, Kay (2003): "On Collective Identity". In: *Protosociology* 18–19, pp. 66–86.
- Manning, Rita C. (1984): "Corporate Responsibility and Corporate Personhood". In: *Journal of Business Ethics* 3 (1), pp. 77–84.
- Mark, Gregory A. (1987): "The Personification of the Business Corporation in American Law". In: *The University of Chicago Law Review* 54(4), pp. 1441–1483.
- Mele, Alfred R. (2000): "Reactive Attitudes, Reactivity, and Omissions". In: *Philosophy and Phenomenological Research* 61(2), pp. 447–452.
- Neuhäuser, Christian (2011): *Unternehmen als moralische Akteure*. Berlin: Suhrkamp.
- Nussbaum, Martha C. (2001): *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.
- Ozar, David T. (1985): "Do Corporations Have Moral Rights?". In: *Journal of Business Ethics* 4(4), pp. 277–281.
- Parfit, Derek (1986): *Reasons and Persons*. Oxford: Oxford University Press.
- Parkinson, Brian/Fischer, Agneta H./Manstead, Antony S. R. (2005): *Emotion in Social Relations. Cultural, Group, and Interpersonal Processes*. New York: Psychology Press.
- Pettit, Philip (2003): "Groups with Minds of their Own". In: F. Schmitt (Ed.): *Socializing Metaphysics*. New York: Rowman & Littlefield, pp. 167–193.
- Prinz, Jesse J. (2004): *Gut Reactions. A Perceptual Theory of Emotions*. Oxford: Oxford University Press.
- Prinz, Jesse J. (2011): "Is Empathy Necessary for Morality?". In: A. Coplan, & P. Goldie (Eds.): *Empathy: Philosophical and Psychological Perspectives*. Oxford: Oxford University Press, pp. 211–229.
- Ranken, Nani L. (1987). "Corporations as Persons: Objections to Goodpaster's 'Principle of Moral Projection'". In: *Journal of Business Ethics* 6(8), pp. 633–637.
- Rorty, Amelie Oksenberg (1978): "Explaining Emotions". In: *The Journal of Philosophy* 75(3), pp. 139–161.
- Rovane, Carol (1998): *The Bounds of Agency. An Essay in Revisionary Metaphysics*. Princeton: Princeton University Press.
- Runciman, David (1997): *Pluralism and the Personality of the State*. Cambridge: Cambridge University Press.
- Salmela, Mikko (2012): "Shared Emotions". In: *Philosophical Explorations* 15(1), pp. 33–46.
- Scheler, Max (1926): *Der Formalismus in der Ethik und die material Wertethik. Neuer Versuch der Grundlegung eines ethischen Personalismus*. Bern: Francke 1980.
- Schmid, Hans Bernhard (2014): "The Feeling of Being a Group: Corporate Emotions and Collective Consciousness". In: C. von Scheve/M. Salmela (Eds.): *Collective Emotions*, Oxford: Oxford University Press, pp. 3–16.

- Scruton, Roger (1989): "Corporate Persons". In: *Proceedings of the Aristotelian Society, Suppl. Vol. 63*, pp. 239–266.
- Sheehy, Paul (2006): *The Reality of Social Groups*. Aldeshot: Ashgate.
- Solomon, Robert C. (1993): *The Passions. Emotions and the Meaning of Life*. Indianapolis: Hackett.
- Steinbock, Anthony J. (2014): *Moral Emotions. Reclaiming the Evidence of the Heart*. Evanston, Ill.: Northwestern University Press.
- Stoll, Mary L. (2005): "Corporate Rights to Free Speech?". In: *Journal of Business Ethics* 58(1–3), pp. 261–269.
- Szanto, Thomas (2012): *Bewusstsein, Intentionalität, und Mental Repräsentation. Husserl und die Analytische Philosophie des Geistes*. Berlin, Boston: de Gruyter.
- Szanto, Thomas (2014): "How to Share a Mind: Reconsidering the Group Mind Thesis". In: *Phenomenology and the Cognitive Sciences* 13(1), pp. 99–120.
- Szanto, Thomas (2016a): "Husserl on Collective Intentionality". In: A. Salice/H. B. Schmid (Eds.): *Social Reality. The Phenomenological Approach*. Dordrecht: Springer.
- Szanto, Thomas (2016b): "Collectivizing Persons and Personifying Collectives: Reassessing Scheler on Group Personhood". In: T. Szanto/D. Moran (Eds.): *The Phenomenology of Sociality: Discovering the 'We'*. London/New York: Routledge.
- Szanto, Thomas (2015): "Collective Emotions, Normativity, and Empathy: A Steinian Proposal". In: *Human Studies*.
- Tollefsen, Deborah P. (2003): "Participant Reactive Attitudes and Collective Responsibility". In: *Philosophical Explorations* 6(3), pp. 218–234.
- Tuomela, Raimo (2007): *The Philosophy of Sociality. The Shared Point of View*. Oxford: Oxford University Press.
- Vargas, Manuel (2013): *Building Better Beings. A Theory of Moral Responsibility*. Oxford: Oxford University Press.
- Wellman, Christopher C. (1999). "Liberalism, Communitarianism, and Group Rights". In: *Law and Philosophy* 18(1), pp. 13–40.
- Wendt, Alexander (2004): "The State as Person in International Theory." *Review of International Studies* 30(2), pp. 289–316.
- Zahavi, Dan (2011): "Empathy and Direct Social Perception: A Phenomenological Proposal". In: *Review of Philosophy and Psychology*, 2(3), pp. 541–558.
- Zahavi, Dan (2014): *Self and Other: Exploring Subjectivity, Empathy, and Shame*. Oxford: Oxford University Press.

Alessandro Salice

# Collective Intentionality and the Collective Person in Max Scheler

**Abstract:** For a good part of the previous century, the notion of a group mind, i.e., the notion of a collective mind distinct from – but not independent of – the minds of the group’s members, was seen as an at best spooky and at worst politically suspect idea. However, in the last thirty years, this idea gained – or maybe regained, as it is argued in this paper – a respectable scientific status: within social ontology one can observe a growing trend which conceives of the notion of groups with minds of their own as an indispensable element of the social world.

The present article discusses one important precursor of contemporary theories about group personhood: Max Scheler’s theory of *Gesamtperson* (or “collective person”). It suggests that there are good reasons to believe that Scheler’s notion of collective person is at least as complex and comprehensive as the one at stake in today’s social ontology, but eventually more radical. To substantiate this claim, the paper is organized into three parts: it starts by illustrating Scheler’s suggested parallel between kinds of groups and forms of sociality. In the second part, it focuses on the idea of the collective person and on its relation to a kind of group Scheler labels “community.” Finally, the third and last part of the paper sheds some light on specific features of collective persons.

## 1 Introduction<sup>1</sup>

For a good part of the previous century, the notion of a group mind, i.e., the notion of a collective mind distinct from – but not independent of – the minds of the group’s members, was seen by philosophers as an at best spooky and at worst politically suspect idea (for a reconstruction – and criticism – of some anti-collectivist views, cf. Pettit 2014). However, in the last thirty years, this idea gained – or maybe regained, as I would like to argue in the present paper – a respectable scientific status: within social ontology one can observe a growing trend which conceives of the notion of groups with minds of their own as an indispensable element of the social world.

---

<sup>1</sup> The author’s work on this article has been supported by a fellowship of the Fritz Thyssen Foundation (Az. 10.13.1.015).

Especially Philip Pettit claims that, in order to secure the possibility of rational decisions at the collective level, one has to assume that there are group minds with beliefs and even desires of its own (Pettit 2003). This view stipulates that, under certain conditions, not only is it possible, but it is also *rational* for the group to hold beliefs that the majority of its members does not hold (this is one of the insights secured by List's and Pettit's recent discussion of the so-called "discursive dilemma," cf. List/Pettit 2011).

Pettit's idea fits finely with what Carole Rovane argues for in her theory of personal identity (Rovane 1997, but cf. Rovane 2014 for some important divergences between the two theories). Rovane claims that personhood is characterized by the commitment one undergoes to reach overall rationality in one's actions and beliefs. In other words, what makes someone a person (i. e., the unity and continuity of one's first person point of view) is his or her striving to achieve overall rationality in his or her beliefs and actions. Still, and this is one of Rovane's points that plays a major role for Pettit's theory, it is not the case that only individuals strive for overall rationality: even groups can do that, and this is what makes certain groups – the groups acting under that specific commitment – *persons* in their own right.

If compared with Pettit's and Rovane's view, Max Scheler's theory of *Gesamt-person* or of collective person does not show a mere terminological affinity with the expressions of "group person" or "group mind." There are good reasons to believe that Scheler's notion of collective person is at least as complex and comprehensive as the one at stake in today's social ontology, but eventually more radical. To substantiate this claim, the paper is organized into three parts: I start by illustrating Scheler's suggested parallel between kinds of groups and forms of togetherness. In the second part, I focus on the idea of the collective person and on its relation to a kind of group Scheler labels "community". Finally, in the third and last part I clarify some interesting features of collective persons.

However, a caveat is in order before starting: an exhaustive account of Scheler's view would have to tackle the two concepts defining a collective person, i. e., such an account would have to say something about what it means to be a *person* and what makes a person a *collective* person. But due to its complexity, Scheler's concept of person deserves its own treatment and I will not be able to address such concept in the context of this paper. That is why in the following I shall leave the idea of person as undefined and concentrate only on those features that make a person a *collective* person.

## 2 Max Scheler: forms of sociality and kinds of groups

Scheler maintains that there is a parallel between different kinds of groups and the different forms of togetherness with others (*Arten des Miteinanderseins*, GW II, p. 529). Call these forms of togetherness also the forms of “sociality” that lived experiences can take: for Scheler, there are three such forms of sociality, one of which is a genuine form of sharing an experience.

*i. Gefühlsansteckung.* The first form of sociality is that of emotional contagion (GW VII, pp. 25 ff). According to Scheler, emotions can spread like viruses among individuals: suppose you enter into a discussion with a sad person, it might be that her sadness contaminates you to the effect that you also become sad – and this without your being aware that her emotion has been transferred to you (GW VII, p. 26). More importantly, there is a sense in which one could say that emotional contagion generally implies illusion, for one believes to be the genuine origin of the emotion, whereas in reality this mental state has been originated in someone else and has passed from mind to mind through contagion (such that only by means of “inferences and causal considerations” can one become aware of the authentic origin of one’s mental state, cf. GW VII, p. 26). Scheler points out that emotional contagion is a watered-down form of identification or *Einsföhlung* (GW VII, p. 29), which is a general psychological mechanism that takes place when one identifies oneself with someone else. Identification plays a crucial role for the constitution of plural subjects and I will come back to this mechanism in section 2.

*ii. Nachfühlen.* One way to understand what Scheler means by ‘Nachfühlen’ is by first looking into sympathy, understood in the sense of *Mitgeföhl* (cf. GW VII, pp. 24 ff). Typically, one feels sympathy when one rejoices in the joy of someone else or commiserates with his or her sorrow. Sympathy, according to Scheler, is a complex phenomenon, for one first has to be aware of the emotion of the other in order to then sympathize with him or her. Although Scheler is not entirely clear about this, it seems that such grasping can occur in different ways. One of them is by means of acts of “vicarious feelings” or *Nachfühlen*, which can be associated with what today goes under the label of “empathy” (cf. Zahavi 2008). It is not immediately evident what Scheler means by this expression, but he illustrates it with cases of so-called presentifications (*Vergegenwärtigungen*). If, for instance, one once saw a landscape and one is now making it present to one’s consciousness, then one has the landscape intuitively in front of oneself – although one is not perceiving it in the proper sense of the term (perception being about the self-given object). Similarly, one can vicariously feel what some-



one else is feeling – this does not mean that one is feeling what the other person is feeling, but only that one is making it intuitively present to oneself (for this example, cf. GW VII, 20).

If compared to contagion, empathy is a state of the subject that generally is not subjected to illusion. That is, if one empathizes with someone, one relates to the other's emotion, and it is clear to the empathizer that it is the other's sorrow or joy emotion and not one's own that one is concerned with.

iii. *Miteinanderfühlen*. Finally, the third form of sociality is that of sharing mental states with another person. This form of sociality can be illustrated by Scheler's well-known example wherein two parents *together* mourn the death of their child (GW VII, pp. 23f). Scheler seems to claim that in this scenario there is only *one* single mental fact that is occurring and not merely two different mental states of the same type (GW VII, p. 24, 48, 75, 252). Co-feeling (*Miteinanderfühlen*) is characterized as feeling a certain emotion *together*. Such notion of sharing or of "Co-" (*Miteinander-*) needs not be confined to collective emotions, though. For instance, cognitive acts and actions can be shared as well, which makes co-feeling only a kind of the more general form of sociality, namely, that of co-experiencing (*Miteinandererleben*, GW II, p. 529f).

Having briefly discussed the three kinds of sociality that lived experiences can exhibit, it is now time to address the parallel Scheler draws between kinds of sociality and kinds of groups. Scheler counts four kinds of groups, namely, masses, communities, societies and collective persons. This is the order in which Scheler introduces and describes them in his *Formalismus*, and such order is not accidental as it is supposed to highlight certain foundational relations between them. However, the first thing that perhaps will grab the reader's attention is that Scheler describes *three* kinds of sociality, but he introduces *four* kinds of groups. Consequently, one might wonder what the missing form of sociality is supposed to be. Before addressing this problem, it could be helpful to first characterize these types of groups and highlight some of their fundamental features.

i. The most primitive kind of group is the mass (cf. GW II, p. 529) – the term "primitive" is used here not in its logical sense, but rather as contrasted to cultivated, designed and/or complex. The mass is characterized by emotional contagion and, more generally, by involuntary imitation of the other. This process of imitation is not restricted to emotions, for it can extend to beliefs and patterns of behavior. Scheler actually goes so far as to describe the entire process of tradition-following as a form of contagion (GW VII, p. 48). Remember that in contagion one generally is not aware of having been contaminated, so one wrongly believes that a certain mental state originates in oneself whereas it has actually originated in someone else. Since, within the mass, the individual ignores having anything in common with the other individuals, i. e., since the individual ignores

*being together* with the others according to this specific form of sociality, no one feels *responsible* for the group behavior that contagion elicits in the crowd's members. At this juncture, it is important to stress that, for Scheler, the notion of responsibility is closely tied to that of solidarity in the sense that, if one is responsible for something together with someone, one is also solidary with him or her. Accordingly, since there is no responsibility between the members of the mass, so is there no solidarity either (GW II, p. 530).

ii. The second kind of group is the community (cf. GW II, pp. 529ff). Communities are bearers of collective intentionality in the genuine sense in which the members co-experience something identical or co-issue the same act. Scheler contends that it is impossible to trace back mental states in the *we*-form to mental states in the *I*-form supplemented by common knowledge (*viz.* the knowledge that everyone has about the fact that everyone has these mental states and that everyone knows that everyone knows it, etc., GW II, p. 530). Not only are these mental states ontologically irreducible, they also constitute a stream of experiences whose subject is a *we* (GW II, p. 530; “Wirerlebnis, Wirbewußtsein” cf. GW VIII, p. 374). The communal *we* is so encompassing that its members primarily live *in* the community and *for* the community.

Section 2 shall tackle how such encompassment has to be understood more precisely, but it is important to first have a look at how collective responsibility works in communities (cf. GW II, p. 530). Two levels can be distinguished here: a group- and a members-level. At the group-level, since a community is a genuine bearer of mental states and actions, the community is fully responsible for them. At the members-level, since all the members co-experience the same state or co-issue the same action, all the members are co-responsible insofar as they are members of the community. Furthermore, in virtue of the link between responsibility and solidarity briefly introduced above, one could also say that the members are solidary with each other insofar as they are co-responsible. This form of solidarity is directed towards the other insofar as the other represents the same community of which one is a member. According to what Scheler calls “representative solidarity [*vertretbare Solidarität*],” the members are solidary with each other primarily because they are members of the community and hence represent the community, but not because they are individual persons in the first place. In a sense, one can find no individual persons within a community, but only representative members of the group. As we will see in the next section, this aspect works differently for the *Gesamtperson*.

iii. The third kind of group covers societies and the dominant form of sociality here is empathy (more precisely, it is a specific form of social cognition that, like *Nachfühlen*, is characterized by a clear self/other differentiation, but that, unlike *Nachfühlen*, is not intuitive, insofar as the other's experience is merely

inferred; on this, cf. GW II, p. 531f.). Remember that the empathizer is the one who adequately discriminates his or her own mental states from others' and who, only in a further step, may (or may not) sympathize with them. Accordingly, a society is composed of individual persons who recognize the other members as individual persons with whom one can enter into social relations and with whom one can join forces to reach goals that for the single individual would be impossible to reach alone. And yet, exactly because of that, i. e., exactly because all the persons involved within a society see each other as distinct individuals and only pursue their (personal) interests, no experiences are shared or co-experienced in the genuine sense at this level. Since no collective experiences occur, so does no collective stream of consciousness take place and, hence, no *we* lives within societies. A further consequence is that there is no collective responsibility among the individuals and, hence, no solidarity either. Rather, *distrust* (*Mißtrauen*) rules among the individuals, and it is such distrust that motivates the establishment of legal institutions whose main aim is to preserve the private interests of the individuals (GW II, p. 532f).

Although communities and societies are radically distinct kinds of groups, Scheler argues that the latter presupposes the former: every society is founded upon a community. This is not to be taken in the sense that the members of a society are at the same time members of a community so that for every society there is a community encompassing the same members. Scheler's point is rather that, in order to enter into a society, individuals must have been or must currently be members of *some* community (GW II, p. 535). This principle is illustrated in different ways, but maybe one particularly clear piece of evidence Scheler adduces is that, in order to exist, every community requires language. For instance, in order to be established, legal institutions need language – but language is not something that can be established in the same way in which legal institutions are established, as this is an element that rather emerges in communities (GW II, p. 535, cf. also Searle 2010 about the uniqueness that distinguishes language from other institutions).

Before addressing the structure and nature of the collective person, it might be useful to briefly recap what has been said so far. I first addressed how Scheler recognizes three forms of sociality: emotional contagion, co-experiencing and empathy. I also highlighted that these three forms of sociality are central to the mental life of the three kinds of groups, namely masses, communities and the societies. Now one might wonder what form of sociality is at the core of collective persons. Maybe not surprisingly, Scheler claims that this is *co-experiencing*: a collective person is a *we* that has a cognitive, a volitive and an affective life of its own. But then, what distinguishes a community from a collective person? The answer to this question seems to be provided by one of Scheler's most in-

sightful contributions to social ontology. To make this clear, one has to take a closer look at his notion of identification or *Einsföhlung*.

### 3 Identification and the collective person

It has been said above that emotional contagion is a watered-down form of identification because, if one is contaminated with an emotion, one believes oneself to be the origin of this emotion, whereas it has actually only been transferred from another subject. By contrast, identification in the full-blown sense takes place when the subject identifies him or herself with someone else in a very specific sense.

Scheler contends that there are two forms of identification: identification can be either idiopathic or heteropathic (GW VII, p. 29). Identification is idiopathic if I identify myself with another subject in the sense that, although I see myself as owner and author of the mental states I am having, I take myself to be someone else. To make use of an example, imagine a young child playing with a doll and pretending to be the mother of the doll. In this case, Scheler argues, the child is identifying with the mother of the doll (GW VII, p. 35). That is, the child sees her experiences as the experiences of herself *qua* mother. The child is aware of being the author or the originator of these mental states, and the child is also aware of being the owner of these mental states: after all, these *are* her experiences. And yet, she *is* the mother of the doll. To illustrate this point, consider the fact that the child would not have thought to have the authority to scold the doll and eventually would not have scolded the doll, had the child not seen herself *as* the mother.

By contrast, identification is heteropathic if the identification goes in the converse direction, i. e., if the other makes me adopting his or her mental states in the sense that I see them as my own. According to Scheler, this is what is going on when a patient under hypnosis adopts the mental states of the hypnotizer. In this case, the subject believes that the hypnotizer's volitions or beliefs are his or her own (GW VII, p. 31). One could hence say that here the sense of ownership is preserved: these volitions or beliefs are seen by the hypnotized as his or her experiences, but the subject does not recognize that the actual author of those beliefs and volitions is the hypnotizer, not him- or herself.

It is now crucial to emphasize that, according to Scheler, heteropathic or idiopathic identification can also have a collective *we* as its target. In other words, one can heteropathically or idiopathically identify either with an individual *or* with a group, and the hypothesis is that collective identification could provide

us with the key to understanding the distinction between communities and collective persons.

It is possible for an *I* to identify with a *we* idiopathically as when an individual identifies with, say, his or her ancestors. In this case, Scheler contends, the individual mystically thinks herself to be the ancestors themselves (GW VII, p. 30). The senses of authorship and ownership are working fine here for, after all, the individual is aware of owning the mental states and of being the author of her states, but she believes herself to be a *we* and – one could add – she believes to have the authority to phrase his or her mental states in the first plural person (on the illegitimacy of this form of authority, cf. Schmid 2014, p. 23).

Heteropathic group-identification works in a slightly different way. Let me first quote Scheler and then comment on his passage:

[...] a man tends, in the first instance, to live more in *others* than in himself; more in the community than in his own individual self. [...] Imbued as [the child] is with ‘family feeling’, his own life is at first almost completely hidden from him. Rapt, as it were, and hypnotized by the ideas and feelings of this concrete environment of his, the only experiences which succeed in crossing the threshold of his inner awareness are those which fit into the sociologically conditioned patterns which form a kind of channel for the stream of his mental environment. Only very slowly does he raise his mental head, as it were, above this stream flooding over it, and find himself as a being who also, at times, has feelings, ideas and tendencies of his own. And this, moreover, only occurs to the extent that the child objectifies the experiences of his environment in which he lives and partakes, and thereby gains detachment from them. The mental content of experience that is virtually absorbed ‘with one’s mother’s milk’ is not the result of a transference of ideas, experienced as something ‘communicated’. For communication entails that we understand the ‘communicated content’ as proceeding from our informant, and that while understanding it we also appreciate its origin in the other person. But this factor is just what is absent in that mode of transference which operates between the individual and his environment. A posited judgment, the expression of an emotional movement etc., is not ‘understood’ in a first stance and experienced as the expression of another *I*, but is rather co-experienced without the ‘co-’ in the ‘co-experiencing’ coming to phenomenal givenness; but this means: it is experienced primarily ‘as’ his own judgment and ‘as’ his own emotional movement. (GW VII, p. 241; my trans.)

It seems that Scheler’s words have to be interpreted in the sense that the *I* can be the bearer of collective thoughts, volitions or emotions, believing that these are his or her thoughts, volitions or emotions, *exclusively*. In these cases, the subject does *not* realize that he or she is only a co-author and a co-owner of these mental states – but still, he or she *is* involved in a collective mental state. This is not something that characterizes only the children’ mental life. Scheler writes:

[...] the vengeful impulses of a member of the family or tribal unit in respect of any insult or injury towards a fellow-member [*Glíedes*] of the same unit is not due to ‘sympathy’ [*Mitge-*

*fühl*] (which already presupposes the givenness of the suffering as the suffering of the other), but to an experiencing of this insult or injury as immediately ‘proper’ [*eigener*] – a phenomenon which is grounded in the fact that the individual initially lives more in the community than it does in himself. (GW VII, p. 242; my trans)

Scheler’s observations about collective identification truly stand out if read against the background of current debate on collective intentionality. For this debate mainly (if not exclusively) focuses on cases where the individual mistakenly frames his or her attitudes as the mental states of a group (Searle 1990). These are scenarios in which, roughly, the individual believes to be a member of a group and to be engaged in collective intentionality, whereas in reality this is not the case. Scheler, by contrast, claims that the most interesting cases in terms of both, commonness and importance for developmental psychology, are those in which collective states are framed by the individual as his or her own (GW III, p. 285). To put this differently, for Scheler the default situation is not the one in which there is an isolated *I* who then intends to join others to form a *we* (and might fail with his or her attempt), rather it is the one in which the *we* encompasses the *I* in a so pervasive way that, as it were, the *I* is phagocytized by – or fully identified with – the *we*.

Scheler goes further by arguing that we are still operating at the level of community even when the individual gains an understanding of such co-authorship and co-ownership – but this is the case only if the understanding at stake is realized *during* and *within* the limits of the co-experiencing itself (GW VII, p. 529). In other words, this is an understanding of being *a mere* element or *an anonymous* member of the community, but it is not – or not yet – an understanding of being *one* individual and *one* specific member of the community among other individual and specific members of the community.

*This understanding*, by contrast, is what characterizes collective persons in contrast to communities. Collective persons are communities, but communities whose members conceive of themselves as individual persons – as persons able to adequately discriminate the proper owner of the experiences they are undergoing. They are persons fully aware of being owners of their own mental states, but also co-owners of collective mental states. In other words, the structure of their attitudes is fully transparent to members of collective persons.

Here it might be helpful to recur to one crucial insight of Scheler’s theory of (individual as well as collective) personhood: personhood is not a property which is conferred by birth, so to say; rather this is a status that needs to be acquired, being *Mündigkeit* (majority, maturity or adulthood) one essential feature of persons. Maturity is understood here as “[...] the possibility to experience [*Erlebenkönnen*] the insight of difference [*Verschiedenheitseinsicht*] between *one’s*

*own and someone else's acts* [...] Or in plain language this means that a man is not of age as long as he simply *co-executes* [*mitvollzieht*] the experiential intentions [*Erlebnisintentionen*] of his environment without first understanding them [...]" (cf. GW II, p. 484; Eng. trans. mod. 479)

Thus, becoming a full-blown person, in a sense, is to free oneself from identification, i.e., it is a cognitive, emotional and practical achievement. Members of communities are not persons from the very outset for they are still subjected to identification (in the sense of *Einsfühlung*<sup>2</sup>) and so they did not have yet developed the ability to conceive of themselves and of the other members as individual persons. Members of communities can free themselves from identification thanks to their ability to exercise empathy, i.e., members of a community can become (aware of their and other individuals' status as) persons in virtue of the fact that they can also become members of societies (i.e., that they can activate empathy). Only after this cognitive passage do they have the possibility of seeing themselves as members of a collective person. (A possibility whose realization can be initiated or favored by historical events, like, for instance, a war, GA IV, pp. 272ff.)

Hence, it is crucial to note that the form of sociality does not change in the community and in the collective person for this remains a *Miteinandererleben*. However, something changes at the level of the subject of collective intentionality. This idea has not to be taken in the sense as if there were two different kinds of plural subjects able to activate collective intentionality or co-experiencing, i.e., communities and collective persons, which both are equipped with different mental powers and abilities. Rather, the idea seems to be that the community acquires specific features that make it a *collective person* in its own right. After all, collective persons are communities, communities that developed into persons (for they are made of mature individuals).

---

**2** In this paper I have used the standard English translation of *Einsfühlung* as "identification," but one might wonder whether this translation is adequate. First, this translation does not capture the affective dimension expressed by the German expression (literally, *Einsfühlung* should be rendered into "the *feeling* of being identical [or of being one]"). Secondly, the technical expression of "identification" or of "group identification" denotes something quite different within social psychology and collective intentionality studies – here, "group identification" rather points to the psychological mechanism which leads one individual thinking of herself as a member of the group (cf. Bacharach 2006), but *not* to the mechanism by means of which the subject misidentifies ownership and authorship of her attitudes.

## 4 Collective persons: Two further features

Based on what has been said so far, we could briefly clarify two further insights of Scheler's view about *Gesamtpersonen*.

The first point is Scheler's counterintuitive claim that *collective* and individual persons are co-originary, i.e., that every person is at the same time an individual *and* a member of a collective person (GW II, pp. 523ff). One way to make sense of this claim is by presupposing that every individual is a member of a community from the very outset. This is because, as we saw, being a member of a *Gesamtperson* actually seems to imply being a member of a community. And individuals always are members of a community, Scheler contends, because individuals come with the possibility of performing social and collective acts; this is a possibility, which is grounded in the essence of a human being (GW II, p. 525). So, for instance, even the lonely stranded Robinson Crusoe would understand himself as a member of a community, although this is a community he would miss in his daily life in the island.

The second point concerns responsibility and solidarity. We said that, in communities, one is co-responsible only to the extent that one co-performs the collective action and one is solidary with the others only insofar as they are members of the community and represent it. Things are different when it comes to collective persons: here the member is not only co-responsible for the other members of the community for the member is also co-responsible for the collective person as well. That is, one is responsible not only for "one's part," but for the entire collective person. This holds for all other members as well and has consequences for the form of ingroup solidarity. Indeed, within collective persons one is not solidary with the other only insofar as the other is a member of the group, but insofar as it is an individual person (GW II, p. 537). Hence, I care about the other not only to the extent that the other is a member of my community, that is, to the extent that his or her actions are contributions to the life of the community. Rather, I care about the other because the entirety of his or her actions matters to me and not because these are relevant to – or are contributions to – the communal life.<sup>3</sup>

---

<sup>3</sup> This has important consequences for the moral assessment of collective persons, on this cf. Szanto (2015).



## Bibliography

- Bacharach, Michael (2006): *Beyond Individual Choice. Teams and Frames in Game Theory*. Natalie Gold/Robert Sugden (eds.). Princeton: Princeton University Press.
- Scheler, Max (1954): “Der Formalismus in der Ethik und die materiale Wertethik. Neuer Versuch der Grundlegung eines ethischen Personalismus”. In: *Max Scheler. Gesammelte Werke*. Vol. II, Bern: Francke, pp. 7–258. [Eng. trans. by Manfred S. Frings/Roger L. Funk (1973): *Formalism in the Ethics and Non-Formal Ethics of Values. A New Attempt Toward the Foundation of an Ethical Personalism*, Evanston: Northwestern University Press.] (GA II)
- Scheler, Max (1972): “Die Idole der Selbsterkenntnis”. In: *Max Scheler. Gesammelte Werke*. Vol. III, Bern: Francke, pp. 213–292. (GA III)
- Scheler, Max (1982): “Der Krieg als Gesamterlebnis”. In: Manfred Frings (Ed.): *Max Scheler. Gesammelte Werke*. Vol. IV, Bern: Francke, pp. 272–282. (GA IV)
- Scheler, Max (1973): “Wesen und Formen der Sympathie”. In: Manfred Frings (Ed.): *Max Scheler. Gesammelte Werke*. Vol. VII, Bern: Francke. Eng. trans. by Peter Heath (1954): *The Nature of Sympathy*. London: Routledge. (GA VII)
- Scheler, Max (1960): “Die Wissensformen und die Gesellschaft”. In: *Max Scheler. Gesammelte Werke*. Vol. VIII, Bern: Francke. (GW VIII)
- List, Christian/Pettit, Philip (2011): *Group Agency. The Possibility, Design, and Status of Corporate Agents*, Oxford: Oxford University Press.
- Pettit, Philip (2003): “Groups with Minds of Their Own”. In: Frederick Schmitt (Ed.): *Socializing Metaphysics*, Lanham: Rowman & Littlefield Publishers, pp. 167–194.
- Pettit, Philip (2014): “Group Agents are Not Expressive, Pragmatic or Theoretical Fictions”. In: *Erkenntnis* 79, pp. 1641–1662.
- Rovane, Carol, A. (1997): *The Bounds of Agency. An Essay in Revisionary Metaphysics*, Princeton: Princeton University Press.
- Rovane, Carol, A. (2014): “Group Agency and Individualism”. In: *Erkenntnis* 79, pp. 1663–1684.
- Schmid, Hans Bernhard (2014): “Plural Self Awareness”. In: *Phenomenology and the Cognitive Sciences* 13, pp. 7–24.
- Searle, John R. (1990): Collective Intentions and Actions. In John R. Searle (2002): *Consciousness and Language*, Cambridge: Cambridge University Press, pp. 90–105.
- Searle, John R. (2010): *Making the Social World. The Structure of Human Civilization*. Oxford: Oxford University Press.
- Szanto, Thomas (2015): “Collectivizing Persons and Personifying Collectives: Reassessing Scheler on Group Personhood”. In: Thomas Szanto/Dermot Moran (Eds.): *The Phenomenology of Sociality: Discovering the “We.”*. London: Routledge, pp. 296–312.
- Zahavi, Dan (2008): “Simulation, Projection and Empathy”. In: *Consciousness and Cognition* 17, pp. 514–522.

Hans Bernhard Schmid

## Being Well Together – Aristotle on Joint Activity and Common Sense

**Abstract:** According to Aristotle, our ultimate purpose in life is to be well, and well-being is in virtuous action over a lifetime. While the concepts of well-being and virtue are subject to heated controversies in the received literature, most interpretations tacitly assume a *distributive* reading of Aristotle's basic claim. In this reading, well-being is what *each of us* wants in *his or her own life*, and it is in *the agent's own* virtuous action over *his or her own* lifetime. A distributive reading can easily accommodate other-regarding and impartial attitudes, as well as the view that no agent can be truly well without the well-being of those with whom he or she lives together. However, no distributive reading can accommodate the view that agents who live closely together in egalitarian relations *participate* in each other's well-being in such a way that one agent's well-being is, in parts, also another's. In his analysis of virtuous friendship, Aristotle points towards such a participatory view. It has been claimed repeatedly in the received literature that any such view has unacceptable consequences, as it extends the subject of well-being in a way that ignores the basic separateness of persons. Taking another person's well-being to be one's own seems to disrespect the other person's own agency. The core claim of this paper is that these passages should be reconstructed as suggesting a collective reading of Aristotle's basic view that complements rather than replaces the distributive view. In this collective reading, what *we jointly* want to do in our *shared life* is to be *well together*, and that being well together is in virtuous *joint activity* over the time of living together. It is argued that Aristotle's participatory conception of well-being does not undermine the participants' own agency as it is in their *common sense* of action that they are unified to a plural subject of well-being. Common sense of joint action is the participants' plural pre-reflective self-awareness of their action *as theirs, collectively*. The extension of the subject of well-being to we-groups broadens our view on how well-being is subjective or first-personal and opens up a plural perspective on the good.

On the Aristotelian account, being well is something that we *do* rather than something that happens to us. It relates to agency in at least two ways. First, well-being (εὐδαιμονία) is our ultimate goal – it is what we pursue, in our activity, for its own sake (EN 1094a). Other goals we pursue with some further purpose in mind. Well-being, in contrast, is whatever we pursue because it is

good in itself. Second, well-being relates to action in that it consists in a *form of action*. Well-being is not some product (κτῆμά τι), but rather some activity (ἐνέργεια τις; EN 1169b 30 f.). Well-being is not a purpose that is outside of activity and towards which activity is directed as its external goal (in the way we engage in brushing our teeth in order to have healthy teeth); rather, well-being is in a special form of activity that has its goal in itself (in the way we engage in listening to music for its own sake). Aristotle's ethics is an examination of the particular form of activity that constitutes well-being.

In the *Nicomachean Ethics*, Aristotle does not ignore that well-being requires external goods (ἐκτὸς ἀγαθὰ; 1099a 31) such as wealth, health, status, and friends. But these goods are goods only in relation to the intrinsic good they make possible which, Aristotle claims, is virtuous living (cf. Cooper 1985). Formally speaking, virtuous living is the exercise of the rational faculties of our soul over a lifetime (1098a 16 f.), and Aristotle's ethical writings contain a catalogue of the single virtues and their mutual relation.

A consequence of the two ways in which well-being is related to agency is that well-being is *subjective*. The subjectivity in question is of the ontological rather than the epistemic kind. Well-being is not whatever you happen to consider as such. At the beginning of the *Nicomachean Ethics*, Aristotle refutes some widely held views of well-being and thus advocates an epistemically objective view; people may *think* that their well-being is in acquiring riches or experiencing lust, but, according to Aristotle, they are wrong. Rather, the subjectivity in question is of the ontological kind. It is a matter of *whose* well-being we are talking about. Well-being is a form of activity, and activity has an agent whose activity the activity in question is. As the form of activity that is well-being is the ultimate goal of the reasonable agent whose activity it is, it follows that the ultimate goal of any reasonable agent is his or her *own* well-being. A modern way to express the ontological subjectivity of well-being is to say that it is *first personal*. This is to say that well-being is what it is *for an agent*, and the relevant agent is he or she whose activity, if it is well-guided, constitutes his or her own well-being.

If it is true that what we ultimately want in our lives is to be well, it thus appears that this can only mean that what each of us ultimately wants in his or her life is his or her own well-being. This seems to be a conceptual consequence of the first-personal nature of well-being – but it sounds problematic. There is something wrong about saying that each of us always strives for his or her own happiness. After all, our lives are interconnected. We are not indifferent concerning other people's well-being. Other people matter to us and while subjective well-being may easily be extended to an inclusive form, it seems that our well-being is shared in a way that seems to be misunderstood even in an extended distributive view. There is a strong intuition that we *participate* in

the well-being of those with whom we share our lives and this intuition seems to be at odds with the way in which well-being is conceived in the Aristotelian view. It seems that there is something fundamentally wrong with this way of conceiving of well-being. Our lives are *shared* lives and Aristotle himself acknowledges this: συζῆν, living together, is what we do, and there seems to be something about living together that is not exhausted by each one living his or her own life. If this is the case, it seems that true well-being cannot be first-personal in the way we have encountered. Something seems wrong about conceiving of the ontological subjectivity of well-being. Well-being is not mine and yours, severally or distributively. It is mine and yours *together*, as a whole in which we take parts. This is not to deny that your well-being and my well-being are distinct. We certainly do not share our *entire* lives – your hobbies may not be mine and our shared life may not extend over our entire lives. But our human way of living together is sharing important parts of our lives. Insofar as our life is shared and in the domain of the activities in which we engage together, jointly, as a team, a couple, a family, or a group, it simply seems wrong to say that all there is to us is that we are several individuals, each one striving for his or her own happiness. True, we *are* several individuals, each one striving for his or her own well-being. But we are not *just* that. There is a sense in which important parts of the striving in question, and the well-being it might constitute, is *ours*, together.

Aristotle himself seems to consider an expansion of the subject of well-being in his analysis of virtuous friendship, especially in chapter 9 of book 9 of the *Nicomachean Ethics*, where he seems to suggest that one's own agency may somehow extend into that of one's friends in such a way that the form of one's friends' activity is indeed part and parcel of one's own well-being. This line has been harshly criticized in the recent literature. Perhaps most prominently, Roger Crisp writes in the introduction of his *Stanford Encyclopedia* entry on well-being:

On Aristotle's view, if you are my friend, then my well-being is closely bound up with yours. It might be tempting, then, to say that 'your' well-being is 'part' of mine, in which case the distinction between what is good for me and what is good for others has broken down. But this temptation should be resisted. Your well-being concerns how well your life goes for you, and we can allow that my well-being depends on yours without introducing the confusing notion that my well-being is constituted by yours. (Crisp 2013)

In this view, an extension of the subject of well-being is simply a “collapse of the very notion of well-being” (Crisp 2013). Other interpreters argue that any such extension of one's own agency into another person's is a “delusion” (Smith-Pangle 2003, p. 152) and that what Aristotle, in his right mind, can only mean is that the “notational egoism” or “formally egocentric doctrine” implied in the ontological

subjectivity of well-being does not include, but indeed implies “practical altruism” (Price 1985, 125f).

The following argues for a different view of Aristotle’s thoughts on an expansion of the subject of well-being – it is neither a “collapse of the very notion of well-being”, nor an illusion of extending one’s own agency into another person’s, nor a conception that remains “formally egocentric”. The key to an adequate understanding of how we participate in the well-being of those with whom we live together is to see that the first-personal nature of well-being does not come in the singular only. The first person has a plural, too, and it is in this plural way that well-being is first-personal for agents who live and act together. The expansion of the subject of well-being that is implied in the participatory view of well-being does not sever the links between well-being and agency insofar as joint action implies a plural subject. In a brief survey of the current literature on joint action, I will argue that a version of the plural subject view is plausible.

I shall proceed as follows: In the first section, the Aristotelian thoughts on a potential expansion of the subject of well-being in his discussion of virtuous friendship will be introduced and some of the critical views to be found in the recent literature will be addressed (1.) The second offers the negative argument that no “formally egocentric” conception of well-being, however altruistic and universal it may be, captures the sense in which we participate in each other’s well-being within a shared life (2.) The third section offers a positive account of the first-personal nature of well-being that extends to the plural, and addresses some issues concerning the relation between singular and plural well-being within a life that is neither ever completely shared nor lived in isolation (3.).

## 1

In his *Stanford Encyclopedia* entry on well-being, Roger Crisp calls it “tempting” to read Aristotle’s views on virtuous friendship in such a way that as my friend, your well-being is “part” of mine (Crisp 2013). As a temptation, this reading should, of course, be resisted in Crisp’s view and he makes clear that he thinks so with a view on the fact that any well-being is always a person’s. It is not quite clear, however, whether or not he thinks that Aristotle did in fact endorse an expansion of the subject of well-being – as he has published his own translation of the *Nicomachean Ethics*, one would expect him to have a view on this question. Reading the relevant passages in book 9, however, Crisp’s somewhat ambiguous assessment appears to be fair. Aristotle here seems to present a view that somehow points towards a straightforward participatory reading of the sharing of well-being among virtuous friends without really going all the way. Of special im-

portance in this regard is the argument in chapter 9. Aristotle here argues for the view that even a person who is virtuous and blessed in such a way that she seems to be rather self-sufficient in her well-being still needs friends. Aristotle approaches this issue from a series of perspectives. One particular line of argument, presented in 1169b, goes through a series of steps which lead up to the core straightforward paradox of virtuous friendship that has bothered most interpreters. The first step is that while virtuous action is good in itself, there is further well-being to be gained from *contemplating* good action. It is a source of the kind of pleasure that contributes to well-being to observe good actions if it is accompanied with an understanding of why they are good. Perhaps it is not too far a shot to capture this thought in the above terminology in the following way: While well-being is ontologically subjective, there is an epistemically objective element to it as well. The epistemic subjectivity in question is not the kind that well-being is whatever is *believed* to be well-being, but rather that the *knowledge* of well-being is part of our well-being. Well-being is not just in acting virtuously, but also in judging that an activity is virtuous. In Aristotle's term, *to see and recognize* (θεωρεῖν) virtuous action is pleasurable (ἡδεῖον), too – and a person that is fully happy cannot miss that pleasure. In the next step, this epistemic recognition is qualified in two ways, leading to the two horns of a dilemma. On the one hand, Aristotle argues that while knowledge of *any* virtuous action is pleasurable, it is most pleasurable for an agent to recognize virtuous action that *belongs to him- or herself* (Aristotle uses the term οἰκεῖον here, not αὐτόν – “domestic” rather than “one's own”). On the other hand, Aristotle argues that our capacity to see and recognize (θεωρεῖν) our own actions is limited. We cannot, Aristotle seems to argue, look at ourselves disaffectedly and “objectify” our own actions sufficiently to take that pleasure-yielding cognitive stance towards them (Aquinas, in his comment to the Nicomachean Ethics [1896], argues that some *privatum affectum* – an emotional bias for ourselves – plays the role of the culprit here). The epistemic stance in question requires some distance and it can be taken in a better way towards one's neighbors than towards oneself, so that their way of acting can be better observed and recognized in that theoretical way than one's own (θεωρεῖν δὲ μᾶλλον τοὺς πέλας δυνάμεθα ἢ ἑαυτοὺς καὶ τὰς ἐκείνων πράξεις ἢ τὰς οἰκείας). From a better position to observe and recognize, Aristotle seems to imply, comes greater pleasure.

Thus it seems that we are caught in a dilemma. On the one hand, the epistemic pleasure is greater in observing one's own virtuous way of acting rather than other people's because the target of observation is more pleasurable. On the other hand, the epistemic pleasure is greater in observing other people's virtuous way of acting rather than one's own, because the observation is better. It seems that we cannot have it both ways. Whether we are contemplating our own

virtuous way of acting or other people's, our pleasure is somewhat impaired. It may thus seem that we cannot reap full pleasure from contemplating virtuous activity because for that, we would need to contemplate an agent that is *us*, and another agent at the same time.

In the final step of the argument, however, Aristotle simply forces the two horns of the dilemma together in what may seem to be an act of conceptual violence. He argues that this paradoxical creature who is both another and oneself is exactly what a virtuous friend *is*: he is, at the same time, ἕτερος, another agent, and αὐτὸς that is, oneself (1169b 7; cf. 1166a 32). The view is not that the other is “another self” *from his or her own perspective*; rather, the other is oneself in that his or her activity is one's own. The paradoxical characterization of the virtuous friend as *alter ego* concludes this line of argument why even a person who is fully virtuous and endowed with all other external goods still needs virtuous friends.

Given the paradoxical nature of the conception, it is perhaps not surprising that Aristotle's view of the virtuous friend as an *alter ego* has met with rather sharp criticism in the received literature. Taking one's friend to be oneself amounts to disrespecting his or her own agency and displacing it with one's own. Many interpreters have observed that Aristotle has developed accounts of an extension of one's own agency into that of another person in the context of his views on the nature of slavery, and of the relation between parent and child (cf., e.g., Price 1985, pp. 103 ff.). Not only is it hard to imagine how these paradigms could serve to capture the nature of participation in egalitarian relations: these conceptions seem to be inconsistent in themselves.

Aristotle's view of the nature of power and domination is extreme in that it entails a straightforward sense in which the slave is not an agent of his or her own, but rather an extension of his or her master's agency (cf. 1140a). In his or her role, the slave does not perform any action of his or her own and therefore does not seem to have any well-being of his or her own in Aristotle's account. This is of course not to say that they “participate” in their master's well-being, because any participation requires that the participants have their own part in what they share in common, which is not the case here, as the only well-being there can possibly be in a master-slave-relationship, if any at all, can only be the masters', and it is the slaves themselves that become “parts” of their masters (cf. Politics 1255b 11). The passages on the extension of the masters' agency into that of their slave oscillate between sociological realism and action-theoretic nonsense. On the one hand, it is perhaps a frighteningly realistic account of the harsh facts of the institution of slavery in which power relations are not agency-regarding (for the concept of agency-regarding relations, cf. Rovane 1998, pp. 74 ff.). On the other hand, it is an action-theoretic view that is de-



cluded in ignoring the slaves' own agential viewpoint even within a slaveholder society. Even within extremely oppressive institutions, any individual's activity, if it instantiates action, instantiates the individual's *own* action, however coerced it may be. This is what Aristotle's view of the slave as a "tool for action" ignores.

A similar delusion seems to be at work in Aristotle's account of virtuous friendship. "Aristotle shows how the friend who is loved as another self is, in some important way, cherished as an extension of oneself, an extension that can tempt one into the delusion that this other really is oneself, and as such is able to help overcome one's limitedness and mortality" (Smith-Pangle 2003, p. 152). Smith-Pangle, as well as many other interpreters, points out that the idea of the "alter ego" makes its first appearance in the *Nicomachean Ethics* in the context of an analysis of the relation of parents and children (1161b). Along these and similar lines, the whole idea of some participation in well-being is a mere delusion, and it is not a particularly nice one because the motive seems to be self-aggrandizement, that is, the extension of one's first-personal perspective over the lives and agencies of others. If this is indeed the motive behind Aristotle's argument that friends are needed, it should certainly not be part of a normative conception of well-being. A plausible conception of shared well-being should not be predicated on such self-aggrandizing delusions.

Thus there is much to say in favor of a non-extensional view of well-being, a conception in which each person pursues his or her own happiness, minding the virtues of her own actions along the distributive reading of Aristotle's basic view. Crisp (2013) argues for this view, claiming that well-being is always "a person's", and that any attempt to argue that "your well-being is 'part' of mine" results in a "collapse of the very notion of well-being". A broadly Aristotelian way of arguing for this view is the following. Aristotle's conception of well-being, as introduced in the first book of the *NE*, is that an agent's well-being is the ultimate goal that reasonable agents aim at. Furthermore, Aristotle argues that the ultimate goal that reasonable agents aim at is not a product of action, but a form of activity. This tight link to agency leads to the view that happiness is not only ontologically subjective, but indeed *first-personal*. If happiness is the form of activity that is the agent's ultimate goal, it seems to be conceptually true that the only happiness towards which an agent can strive is the agent's own, because it is only his or her own action that an agent can want to bring about in such a way that the action is not a product of what he or she does, but in the doing itself. This is what Price (1985) refers to as the "formal egocentricity" or the "notational egoism" of Aristotle's account.

Imagine what it may mean to aim at another person's happiness from this perspective. *Ex hypothesi*, another person's well-being is a sustained form of his or her activity. Now it is certainly possible to influence people in such a



way that the form of their activity changes, and in ordinary language, we would probably say that in this case, you're making another person happier. But again, it should be remembered that in the Aristotelian view, well-being is an activity (ἐνέργειά τις) rather than some property (κτῆμά τι; 1169b 30–31). The good that is well-being is a form of the activity itself rather than its product. Thus “making somebody happy” can only mean providing the “external goods” that enable the other to be active in the way that constitutes his or her own well-being. As the well-being you're helping to bring about is in the form of the other person's activity, it is something that *the other person does*, rather than a form of your own activity. Thus in assisting other people in such a way as to make them act in a better way, your acting is of the kind that *contributes* to other people's well-being and, maybe to some degree “makes the other person act”, but insofar as this assisting activity is a part of anyone's well-being, the only possible candidate well-being is your own.

It is thus a conceptual truth, in this conception, that well-being is always each agent's own, simply because it is his or her own agency only that each agent exerts. Even a person who has great powers of command over others, and is thus in some control of what they do, does only the *commanding* rather than the *doings* of his or her subordinates. Each agent does his own doing, even where agents are under other agents' power. One might call this minimal agential autonomy: Wherever an agent acts, the action in question is his or her own. In a conception of well-being according to which well-being is a form rather than a product of action, this minimal conception of agential autonomy seems to force a distributive reading: As each does his or her own doing, each one's well-being is his or her own.

## 2

Whether or not it is something like this first-personal nature of well-being Crisp has in mind when he says that the “temptation” to say that we can literally participate in each other's well-being should be resisted, he goes on emphasizing that any such view should not be taken in the sense of atomistic independence or narrow egoism. This is certainly right and even though it does not take us as far as Crisp hopes, the ways in which the distributive reading of well-being can accommodate mutual interdependence, altruism, and non-particularism deserve to be mentioned.

Your well-being concerns how well your life goes for you, and we can allow that my well-being depends on yours without introducing the confusing notion that my well-being is

constituted by yours. There are signs in Aristotelian thought of an expansion of the subject or owner of well-being. A friend is ‘another self’, so that what benefits my friend benefits me. But this should be taken either as a metaphorical expression of the dependence claim, or as an identity claim which does not threaten the notion of well-being: if you really are the same person as I am, then of course what is good for you will be what is good for me, since there is no longer any metaphysically significant distinction between you and me. (Crisp 2013)

The non-threatening version of the identity claim is the one in which “you are the same person as I am” simply means “you are the same *kind of person*”. It is qualitative or type identity that Crisp advocates here, not numerical or token identity. It may not initially seem convincing, however, that my being of the same *type* of person as you solves the problem of the first-personal nature of well-being. Just the fact that I like to do the same type of activities that you like does not *per se* bring us together in any friendly way – in a world where resources are scarce, conflict easily results from people having preferences of the same kind. The type identity claim clearly needs to be specified in order to capture something of the spirit of “being well together”.

There are several steps to take. First and perhaps most basically, the “dependence claim” Crisp mentions has to be taken into account. Ways in which “my well-being is closely bound up with yours” can obviously be accounted for without any expansion of the subject of well-being, and without blurring the neat distributive distinction between how well your life goes for you and how well my life goes for me. Aristotle recognizes that happiness requires suitable conditions. Here, the issue of “external goods” is important. Aristotle lists having nice children, some wealth, and being reasonably good-looking as paradigmatic examples in this category, but it is perhaps not too much of a stretch to claim that in order for a person to be truly happy, it is necessary that his or her loved ones be reasonably happy, too. After all, it is difficult to imagine that a virtuous person be completely happy if everybody around her is just plain miserable, and given the nature of Aristotle’s inquiry, one likes to think that he must have said so somewhere. There is *some* sense in which well-being obviously involves sharing here. In the *Eudemian Ethics*, Aristotle emphasizes the *sympathetic* nature of close friendship, arguing that we want to share even our friends’ pains, and if it is not possible to share the *same* token experience of thirst or hunger, then we want to experience at least “what is closest to it” (EE 1240a 39). It seems plausible that this extends beyond single episodes and stretches to our friends’ whole well- or unwell-being. This kind of sharing – sympathy – does not challenge the first personal conception of well-being in any way. Even if it is true that well-being involves so much sympathy that I can be truly well only if you are, too, or if it is true that I can be truly well only if everybody

else is, it remains a fact that in spite of this dependence, my well-being is not yours and yours is not mine. A condition of my and your well-being may be that we are equally well and it may be exactly the same kind of activity that makes each of us happy. But even if we're both equally well and if it is the same sort of activities that constitute our well-being, your happiness is yours and my happiness is mine.

The next step is that in addition to dependence and qualitative identity, it has to be recognized that Aristotle's first-personal account of well-being, even in its most straightforwardly "egocentric" reading, certainly does not entail crude egoism. The first-personal way in which Aristotle spells out the ontological subjectivity of well-being has to be distinguished clearly from the view that agents are self-centered. It is true that Aristotle does seem to ascribe a great deal of egoism to agents (cf. Kahn 1981): "Each one seems to love his own good", Aristotle says (1155b 23), and again: "everyone wants good things most of all for himself" (1159a 11–12). But the qualification in the first quote already makes clear that egoism is not a conceptual truth about agency, and it seems that Aristotle is talking about the average unhappy individual here rather than the virtuous person who is truly well. Aristotle's analysis makes it quite plain that a lot of my close ones, and indeed society at large, will profit a great deal from my true well-being, and that they will do so not only as an unintended side-effect of my conducting my life well, but as part of what it is for me to lead a good life. This is obvious from the fact that generosity is a paragon feature of the good life, as Aristotle argues at length in the catalogue of virtues, and fairness plays a key role, too. Aristotle's conception of happiness, as well as any plausible conception, states that a good life involves finding a balance between the pursuit of one's own interest and an engagement for the interests of others. On this line, Aristotle says of the "virtuous man" that his "conduct is often guided by the interests of his friends and of his country" and that "he will if necessary lay down his life on their behalf, thus ". This, however, is not against the reasonable self-interest of people, as they thereby "chose great nobility for themselves" (1169a). Thus the claim that well-being is first-personal does not entail that well-being should be pursued in a self-centered frame of mind or that the only interests a reasonable agent takes into account are his or her own. Subject and beneficiary of well-being are two separate issues. The subject is the first person, but the circle of beneficiaries is much wider; the question of whose well-being is at stake does not answer the question of who profits from one's well-being.

This leads us to the last (and perhaps most important) step. In Aristotle's view, well-being is about proper guidance by reason, entailing values such as justice. And such values are universal rather than particular in nature. This is not to deny that we do tend to endorse self-serving conceptions of justice; but

to the degree to which we do, we're not truly just. Justice is what it is not just *for me*, but *for anyone*. Insofar as well-being is in a life that is guided by reason, it is, in a way, a *universal* life; it is guided by the parts of our souls in which there is, indeed, "no metaphysically significant distinction between you and me", as Crisp puts it. The first-personal character of well-being does thus not entail particularism, but rather entails a great deal of universalism. This is the point at which, in spite of all the differences, some interpreters have found a connection between Aristotle's conception of virtue and Kantian thoughts on duty (e.g., Sherman 1997). The virtuous life is universal in its being guided by reason and universal values.

It is time to take stock and see how far these steps have taken us. Has our problem with the first-personal distributive reading been solved? Crisp suggests that if the mutual interdependence between our well-being, the difference between the ontological subject and the beneficiary of well-being, as well as the universalism of Aristotelian well-being is properly understood, nothing remains unaccounted for in the view that well-being is about each of us pursuing his or her own well-being. But this is not so. Even if it is true that my well-being presupposes yours, and that it involves so much sympathy that I can be truly happy only if you are, if it is true that I can be truly well only if everybody else is, and that it is in the same kind of activity that each of us finds his or her happiness, the plain fact remains that my happiness cannot be part of yours and your happiness cannot be part of mine. And this seems to be a serious limitation. In close friendship, that is, in an intimately shared life among equals, the well-being is shared in a different and more straightforward sense than the "non-threatening" sense advocated by Crisp. It is not just that friends tend to be equally well or unwell and that their well-being is mutually interdependent and in the same kind of activities. Rather, they participate in each other's well-being. If we live closely together, it is certainly the case that each of us pursues his or her own well-being and that our individual well-beings are closely interrelated. But this is not all there is to a shared life. In intimate relations, the striving for well-being is not something each member pursues by him- or herself, in a more or less generous, sympathetic, and universal fashion. Rather, it is something we *do together* and whatever well-being is in activity that is not *mine and yours, severally*, but *ours, together*, it seems clear that the well-being in question is ours in such a way that we literally participate in each other's well-being. This is the sense excluded by Crisp's version of the identity claim, but it is a sense that seems plausible and indeed constitutive of the kind of well-being we may enjoy in a life that is truly shared with others. Are there ways to account for this without thereby causing a "collapse of the very notion of well-being"?

## 3

That *joint activity* is a key to understanding the participatory nature of well-being that Aristotle seems to advocate in his conception of virtuous friendship has not gone unnoticed in the received literature (cf., e.g., Cooper 1977; Price 1989, pp. 103–130; Sherman 1997, chap. 5; Kosman 2004). To focus on just some of these examples, Anthony W. Price’s chapter on “perfect friendship” (1989, pp. 103–130) is guided by the intuition that the kind of participation in each other’s well-being that seems to be implied in Aristotle’s account cannot be captured adequately in any other way as in an analysis of what it means to cooperate or act jointly, together, and he undertakes a thorough examination of the conceptual tools Aristotle offers to understand the kind of interrelation between cooperating agents and their participation in joint activity. A second example out of many is Nancy Sherman’s *Making a Necessity of Virtue* (1997), especially the chapter entitled “the shared voyage”. “Doing things together” and “friendship and shared activity” are the key topics here and Sherman again focuses on the passages that support the reading of living together in terms of joint activity.

Aristotle characterizes living together (συζῆν) as a “sharing of words and deeds” (λόγων καὶ πραγμάτων κοινωνεῖν; NE 1126b 11). It is true, though, that Aristotle does not mention joint action in other places at which the issue of συζῆν comes up. In 1270b, “κοινωνεῖν λόγων καὶ διανοίας”, “the sharing of words and thoughts (or considerations)” is the expression Aristotle uses, and in the *Eudemian Ethics*, he characterizes the living together again cognitively rather than practically, this time as joint perception and common knowledge (or common acquaintance, or common understanding; συναισθάνεσθαι καὶ συγνωρίζειν, cf. EE 1244b 26–27). The fact that Aristotle’s focus is on the sharing of perception, knowledge, logos, and thinking, however, does not mean that it is only here and not in actual action that the sharing (or “communing”; κοινωνεῖν) takes place in living together. Rather, he focuses on the rational infrastructure of the kind of activity that is “with account” or according to (the right) reason [κατὰ (τὸν ὀρθόν) λόγον] for the following two reasons: It is only this particular kind of action that comes into question of a shared life that is a good life, as the good life is a reasonable life, and the sharing of activity that is integrated at the level of the reasoning is shared in a different sense than spontaneous joint activity, or shared activity that is based on individual reasons and in which each participant ultimately pursues an individual agenda. The idea is that for our shared life to be a good life, our shared activities must issue from a joint rational perspective, and it is in virtue of this perspective that the activity in question is unified.

Given this widely shared interpretation in the received literature, the task ahead is to show that the interpretation of virtuous friendship as joint virtuous activity can accommodate a straightforward participatory reading of well-being that is free of the paradox and the unacceptable consequences encountered above, especially that it is non-delusional and does not undermine basic agency-regard and that involves an expansion of the subject of well-being that does not end up in a paradox and a “collapse of the very notion of well-being”, as Crisp assumes.

Before coming to these issues, let us first retrace our steps to see how the apparent paradox of the participatory reading of well-being emerged. In the Aristotelian view, well-being is tied to action in that it is our ultimate aim, and it is a particular form of activity rather than a product of action. The only activity which we can “do” in a way that is not a product of our activity is *our own* (the *own agency-condition*), this leads to a first-personal account of well-being, according to which the kind of well-being at which each agent aims is his or her own (sect. 1 above). This does not rule out altruistic conceptions of one’s own well-being or a universalist orientation, but it does seem to rule out a participatory conception of well-being in which well-being is shared in a straightforward sense, as we cannot seem to conceive of another person’s action as our own without undermining the other’s own agency (sect. 2 above). The *own agency-condition* and the agency-regard it implies seem to be incompatible with the view that the other’s action is somehow one’s own. If this is true, it follows that you pursue your happiness and I pursue mine and that all there is about sharing well-being is mutual dependence, altruism, and a non-particularist conception of well-being, without any straightforward sharing of well-being between us.

Thinking about virtuous friendship in terms of *joint activity*, however, suggests that the way in which we participate in each other’s well-being may not be a matter of me doing your individual actions, which violates the *own agency-condition*, but rather a matter of *us* acting *together* in the sense that there is *one* activity to which each of us contributes. If we act together, there is one token action that we perform together, but this is not to say that your agency is somehow subsumed under mine, as you contribute your parts while I contribute mine. This suggests a conception of well-being that is both first-personal and agency-regarding, and participatory in a straightforward sense. Joint agency respects the *own agency-condition* because it does not suggest that I do your contributions to the joint activity or that you do mine. As far as joint agency is virtuous however, it constitutes a well-being that is not yours and mine, severally or distributively, but *ours*, collectively.

Thinking further along these lines, the next issue to take up would be an analysis of what exactly virtuous joint action is, and how it relates to individual

virtuous action. Are there virtues that are *specific* for joint activity? A first candidate to consider may perhaps be cooperative-mindedness, that is, the habitualized disposition to engage in the kind of joint activity that is directed towards the common good. To be cooperative-minded is to see and relate to one's partners as competent participants in reasoning and action. It would be interesting to see how exactly the kind of trust and mutual reliance involved in cooperation relates to the kind of self-confidence and self-reliance – trust in *one's own* abilities and judgment – advocated in Aristotle's accounts of the virtues of magnanimity (or “big-souledness”, μεγαλοψυχία).

This would be the topic for another paper. The current aim is a more modest one: It is to point out the flaw in the line of argument that lead to the view that the first-personal nature of well-being cannot be reconciled with the intuition that well-being can be shared in a straightforward sense. This flaw is in the tacit assumption that the first-personal nature of well-being is limited to the *singular*. This is a mistake because there is the first person *plural*, too. The way in which being well together is first personal is in the plural. Participatory well-being is tied to joint action in the same way individual well-being is tied to individual action. Being well together is what we, together, ultimately want and it is in the virtuous form of our joint activity. Such well-being does not imply the delusion that my agency somehow extends into yours and thereby displaces your agency, but it is based on the insight that agency can be joint in such a way that both of us participate in a common endeavor. If this endeavor is virtuous, it constitutes well-being that is not mine or thine, but *ours*.

Thus the “extension in the subject of well-being” at stake here is not the displacement of the second person by the first person, but the move from the first person singular to the first person plural.

It may not initially seem convincing, however, that the mere fact that individuals act jointly somehow involves the “extension of the subject of well-being”. If the subject of well-being is “us, together” in a way that implies a whole that contains us as parts, it seems that the subject of the virtuous joint activity must be plural. Looking at the recent literature on joint action, such a claim seems to receive only weak support. Most authors agree that joint actions are collectively intentional, but only few authors claim that collective intentionality has a plural subject in the sense of there being a group who “has” the intention in question (cf. Schweikard/Schmid 2013). Most authors answer the question of what's collective about collective intentionality by pointing towards the *content* (e.g., Bratman 2014) or the *mode* (e.g., Searle 2010 and Tuomela 2013) of the intentionality in question. For a long time, Margaret Gilbert (1989) has been unique in pushing a plural subject account of collective intentionality, and her account has been met with serious objections. Instead of entering this extended debate – out of



which, in my view, a revised version of the plural subject account emerges victorious –, let us again look at Aristotle and see how a plural subject account of joint activity (and thereby of participatory well-being) could be construed with the action-theoretic tools he provides. There may be a bit of a lacuna to be filled here. As seen above, the existing literature on Aristotle's virtuous friendship as an account of joint action has largely focused on the claim that what is joint about joint action comes from the *joint reasoning* from which the action issues. Yet this only pushes the question further back. After all, reasoning is just another activity. What is the feature in virtue of which *this* activity is joint, that is, a unified activity of reasoning, rather than a case of two or more individual reasonings that are interlinked in such a way that each comes to the same conclusion? What is the feature in virtue of which the conclusion is *one* that is reached *together* rather than an interlinked aggregate of individual conclusions?

It is certainly true that not only activities that issue from deliberation can be joint activities. Just as there are spontaneous individual actions that do not issue from a prior intention (cf. Bratman 1987, pp. 126–127; Velleman 2007), there are cases of spontaneous joint action such as an improvised Jazz jam session (cf., e.g., Tuomela 2007, p. 274). It may not even seem altogether convincing that such spontaneous activities should be somehow less important to our well-being – individually or collective – than fully reasoned choices, as Aristotle clearly claims. More importantly, however, even if one assumes that only reasoned activities can be joint in a way that is relevant for being well together, the question of what exactly is joint in joint activity, is not to be answered simply by pointing out that kind of activity. So the crucial question remains open: what is the feature in virtue of which some activities are joint activities? What would an Aristotelian account of collective intentionality look like?

Looking at the decisive chapter 9 of book IX of the *Nicomachean Ethics* more closely, a feature hits the eye which has not been taken sufficiently into account in the received literature (an exception is Hitz 2011). Aristotle devotes large parts of this chapter to a discussion of a theme which is recurrent in his work: the idea that – liberally translated – human activity is typically *conscious* activity; there is a sort of awareness – the term Aristotle uses is αἰσθησις here – of our activities. “One who sees is conscious that he sees, one who hears that he hears, one who walks that he walks, and similarly for all the other human activities there is a faculty that is conscious of their exercise” (1169b). It is, Aristotle then continues to argue, essentially in virtue of the *self-awareness* that virtuous action constitutes well-being, suggesting that it is self-awareness of virtuous activity that makes existence desirable. The next step in the argument makes the crucial move. Earlier in book 9, Aristotle has pointed out several ways in which an agent's relation to his or her friend parallels, or is identical with, the relation



of an agent to himself, culminating in the parallel between individual self-love and love of friends in virtuous friendship. In chapter 9, Aristotle seems to spell this out in terms of self-knowledge or self-awareness.

To my knowledge, Zena Hitz' interpretation of Aristotle's conception of friendship is unique in that she focuses on the relation between friendship and self-knowledge. In her thorough and careful examination, Hitz points out that it is in the role of self-knowledge of one's friends' action that a solution to the problems that have been diagnosed in previous interpretations can be found. The way in which one's friends' actions are "somehow one's own" is not in some delusional extension of one's own agency, but in a sort of knowledge of the activities in question. But of course, the paradox reappears immediately: how can an agent be self-aware of another agent? It is important to take a closer look at the decisive passage here. The way Aristotle puts his point across is not by claiming that an agent has *the same* αἴσθησις of his or her friend's action or existence as of her own. Rather, Aristotle says that an agent ought to have αἴσθησις of his or her friend's virtuous action *together* with the friend (1170b). The word Aristotle uses here is συναίσθάνεσθαι, "being aware together", meaning that the αἴσθησις in question is a συναίσθησις – in this latter, nominal form, the term only seems to occur once in the *corpus Aristotelicum*. In EE 1245b, Aristotle emphasizes that while it be desirable to extend the sharing of αἴσθησις to many, it can be actualized only with the few with whom one actually lives together; the sharing of one's life determine the limits of that "community of consciousness". Yet it is clear that in spite of these tight social limitations placed on the actual sharing of self-knowledge, Aristotle's remarks on the role of συναίσθάνεσθαι and συναίσθησις prepare the concept of the "shared perception", the "shared sense" (κοινή αἴσθησις) placed at the heart of human sociality in book 1 of the *Politics* (P 1253a).

Aquinas translated Aristotle's κοινή αἴσθησις with "sensus communis" and, over the course of a long and often reconstructed history, this has turned into the "common sense" many social and political philosophers have assigned a core place in their thought, usually giving it the meaning of the basic mental capacities, dispositions, and attitudes people can usually be expected to have and that therefore provides a non-idiosyncratic and impartial standard of judgment. If the interpretation given above is correct, however, Aristotle's conception of the common sense provides a deeper reason why any such common sense should be distributively general. Each of us has common sense because we have common sense *together*, collectively, in that common sense is awareness of the way in which we live and act together.

The later history of common sense has covered up the collective root of common sense. But the difference between a merely distributive understanding of

common sense and an understanding of common sense as collective does matter. In the exclusively distributive reading, the sense in question just happens to be common in something like the way in which each human usually happens to have one nose and two ears. It is only in the reading that extends to the collective dimension that common sense is transparent as the basic feature of the kind of activity that is our shared life. Each of us has common sense for him- or herself *because* we have common sense *together*. Seeing common sense as something in which each of us participates rather than just as something each of us instantiates for him- or herself opens up a perspective on how common sense is *contingent* rather than something like an unchangeable “given” of our social life, as it seems to be seen in conservative views. If common sense is seen as our awareness of the way in which we live together, it is obvious that our common sense is a feature of what we do, jointly, as societies. And this is up to *us, together*.

This brings us to a final point, which concerns the relation between the συναίσθάνεσθαι in virtue of which our activity in our shared lives is a joint activity and the θεωρεῖν of virtuous action that according to NE IX 9 adds to our well-being (cf. above sect. 1). Aristotle claims that in θεωρεῖν, in seeing and recognizing virtuous action we experience a pleasure that is essential for a life that is fully happy. This θεωρεῖν is obviously not the αἴσθησις we have encountered in this section and that is part and parcel of any conscious activity (in the way that any hearing involves the awareness of hearing) – even though both attitudes are self-referential and have the same target, that is, the activity. It is tempting and perhaps not too much of a stretch to reconstruct the difference between these two self-referential attitudes as the difference between pre-reflective, non-thematic, non-objectifying self-awareness on the one hand, and reflective self-knowledge on the other. If this is the case, the kind of θεωρεῖν that makes us fully happy that Aristotle is looking for in NE IX 9 would be reflective self-knowledge of shared activity: making explicit and thematic the pre-reflective συναίσθησις that marks our shared activity.

Here, I claim, is the key to an understanding of how well-being can be participatory without displacing the other as an agent. The way in which agents can be well together in a participatory sense is by having self-knowledge of what they are jointly self-aware of, that is, their shared activity. The θεωρεῖν in question is the knowing that whatever each of us is doing in living together (συζῆν) is unified by a strong common sense. Living well together is conscious activity, and it is not that I am conscious of what I do and have observational or inferential knowledge of what you are doing, knowing that *mutatis mutandis*, the same is true for you and that there is some structure of common knowledge between us. Rather, the consciousness in virtue of which our shared life is *ours, collectively*, rather than yours and mine, distributively, is *plural* consciousness, or

συναΐσθησις, that is, αἰσθησις that is *ours* rather than yours and mine. The consciousness in question is plural pre-reflective self-awareness (Schmid 2013), and in the right way of seeing and recognizing that activity (θεωρεῖν), our activity is not only to be seen as a distribution of individual activities, but it should be grasped from its collective ground. Being well together is acting virtuously together and knowing it.

This step from the singular to the plural is how Aristotle's account of participatory well-being involves an "extension of the subject of well-being". It is not delusional, it does not imply a self-aggrandizing view of the extension of one's power into the domain of what is up to others, and it is certainly not "a collapse of the very notion of well-being", as Crisp claims. Rather, it adequately describes how agency is marked by a sense of self, and how that sense of self is plural where agents live closely together and engage in joint activities. You do not thereby become "I" and "I" am not you. Yet you and I are *us*, and as such, we can be well together in such a way as to participate in a well-being that is not just an aggregate or distribution of individual well-beings, but a collective well-being. It is in this sense that it is not only true that what each of us wants in his or her own life is to be well, but that what *we jointly* want to do in our *shared life* is to be *well together*, too. And it is in this sense that your well-being is not only a prerequisite or a product of my virtuous activity, but something in which we participate insofar as we are living together rather than leading mutually interdependent, but separate lives.

## Bibliography

- Bratman, Michael E. (1987): *Intentions, Plans, and Practical Reasons*. Cambridge: Harvard University Press.
- Cooper, John M. (1977): "Friendship and the Good in Aristotle". In: *The Philosophical Review* 86, pp. 290–315.
- Cooper, John M. (1985): "Aristotle on the Goods of Fortune". In: *The Philosophical Review* 94, No. 2, pp. 173–196.
- Crisp, Roger (2013): "Well-Being". In: Edward Zalta (Ed.): *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/well-being/>, visited on 25 August 2015.
- Hitz, Zena (2011): "Aristotle on Self-Knowledge and Friendship". In: *Philosopher's Imprint* 11, pp. 1–28.
- Kahn, Charles K. (1981): "Aristotle and Altruism". In: *Mind* 90/375, pp. 20–40.
- Kosman, Aryeh (2004): "Aristotle on the Desirability of Friends". In: *Ancient Philosophy* 24, pp. 135–154.
- Price, Anthony W. (1989): *Love and Friendship in Plato and Aristotle*. Oxford: Oxford University Press.

- Rovane, Carol (1998): *The Bounds of Agency. An Essay in Revisionary Metaphysics*. Princeton: Princeton University Press.
- Schmid, Hans Bernhard (2013): “Plural Self-Awareness”. In: *Phenomenology and the Cognitive Sciences* 13, pp. 7–24.
- Schweikard, David P./Schmid Hans Bernhard (2013): “Collective Intentionality”. In: Edward Zalta (Ed.): *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/collective-intentionality/>, visited on 25 August 2015.
- Sherman, Nancy (1997): *Making a Necessity of Virtue. Aristotle and Kant on Virtue*. Cambridge: Cambridge University Press.
- Smith-Pangle, Lorraine (2003): *Aristotle and the Philosophy of Friendship*. Cambridge: Cambridge University Press.
- Tuomela, Raimo (2007): *The Philosophy of Sociality. The Shared Point of View*. Oxford: Oxford University Press.
- Velleman, David J. (2007): “What Good is a Will?”. In: Anton Leist (Ed.): *Action in Context*. Berlin/New York: De Gruyter, pp. 193–215.



Michela Summa

# Pretence and the Inner. Reflections on Expressiveness and the Experience of Self and Other\*

**Abstract:** This paper investigates the relation between pretence, social experience, and the knowledge of self and of other. Traditionally, sceptical arguments concerning both the existence and the knowledge of other minds have been developed on the basis of the remarks concerning the possibility of pretence, and particularly of deceptive pretence. Combining Wittgenstein's approach to pretence, as it is particularly developed in *Last Writings on the Philosophy of Psychology*, with a phenomenologically grounded conception of pretence and intersubjective experience, I aim at circumscribing the meaning of the question concerning our possible knowledge of ourselves and others.

Departing from a critical assessment of the implications of the sceptical arguments based on pretence, I argue that the concept of 'knowledge' underlying such sceptical challenges is misleading. Subsequently, I develop a phenomenological inquiry into the relation between experiencing and expressing, and show how the results of such an inquiry fruitfully complement Wittgenstein's and Cavell's observations about "knowledge" and 'acknowledgement'. I finally return to pretence and its implications for social experience and understanding. I argue that the revised conception of knowledge of self and other developed in the previous sections allows us to dismiss the understanding of pretend actions and behaviour as something that undermines the possibility of social understanding. On the contrary, such a conception of knowledge of self and other brings to the fore the constitutive social nature of pretence.

The possibility of pretence, and correlatively of being deceived by others' behaviour, is one of the arguments brought about while formulating both epistemological and ontological challenges concerning other minds.<sup>1</sup> According to the former, we can

---

\* I wish to thank Andrea Altobrando, Monika Dullstein, and Karl Mertens for their insightful comments on a preliminary version of this paper.

**1** It should be noticed right now that the meaning of the English word "pretence" is not univocal. In particular, a distinction should be made between non-deceptive pretence, in which there is a mutual understanding among subjects concerning the fictional nature of a situation or concerning the fact that one is playing a role (e.g., in pretend play), and deceptive pretence, in German *Verstellen* or *Heucheln*, in which someone is feigning, hiding something, or dissim-

never be certain about what others are experiencing, whereas according to the latter we can never be certain about the very existence of others as experiencing beings. In his later writings, and particularly in *Last Writings on the Philosophy of Psychology*, Wittgenstein focuses on such sceptical challenges within the framework of a reassessment of the distinction between the inner and the outer, which also implies a reassessment of our experience of ourselves and others. As some of Wittgenstein's remarks reveal, both the concept of pretence and its relation to inner experience and genuine expression are far from being clear.

The aim of this paper is to develop a phenomenological account of the relation between pretence and our "knowledge" of ourselves and of others. I will argue that the assumptions underlying the understanding of pretence in the just mentioned sceptical remarks are related to a problematic view of what expression is and of what the privileged access to our own experiences entails. On the one hand, outer expressions are considered to be decoupled from experience and to point at what happens in inner psychic life. On the other hand, our privileged access to our experiences is considered only on the basis of the idea of certainty. In what follows, I shall question both views. In the first section, I critically address the implications of the sceptical arguments based on pretence. In the second section, I concentrate on the relation between experiencing and expressing. This is done by discussing three key-points: the asymmetry between self- and other-experience, the distinction of two different meanings of expression, and the qualification of the concept of knowledge when applied to both our own and the other's experiences or mental states. Thereby, I particularly refer to Wittgenstein's and Cavell's remarks and show in what sense they can complement a phenomenological understanding of experience and expression. Finally, in the third section, I pursue this line of thought in relation to pretence. Here, I discuss the shortcomings of considering pretence as based on already given and fully determined mental states, and develop an interpretation of pretence that, instead of considering it as something that undermines intersubjectivity and social understanding, is rather grounded on both.

---

ulating. In this paper, I shall only concentrate on deceptive pretence. A conceptual analysis of the different meanings of the concept of pretence is developed by Mercolli (2012). For a phenomenological inquiry into non-deceptive pretence, see Summa (2016). On the concept of deceptive pretence, see also Austin (1958) and Anscombe (1958), who debate on the meaning of the concept of pretence as intentional activity aimed at disguising or dissimulating. The discussion between Austin and Anscombe does not particularly touch the implications of pretence for the problem of other minds, which will be the central topic of this paper.

# 1 Pretence and the scepticism concerning other minds

As I mentioned, the sceptical challenge based on pretence can assume two forms: epistemological and ontological.<sup>2</sup>

The epistemological challenge is based on the assumption that we do not have any direct access to the other's inner states; we can only infer them on the basis of what we can perceive, namely outer behaviour. This challenge concerns the possibility of identifying sufficient outer criteria (behaviour, actions, expressions, and language) that would allow us to recognise with certainty the psychic (i. e., inner) states of the other (cf. Witherspoon 2011). The possibility of pretence is assumed as an argument supporting this challenge. Indeed, since one can express, for instance, pain also without being in pain, the perception of the expression does not give us any legitimate ground to say with certainty whether the other is in pain or not. Generalising this remark, the sceptical challenge goes as follows: assumed that others sometimes deceive us, and that they can deceive us in principle, we can never be certain of whether we grasp their genuine states of mind at all. These could, in principle, always remain hidden.

The ontological challenge concerns the very existence of others as minded beings. Not only can I be deceived about the correlation between the other's outer behaviour and his/her inner states, but also about the very existence of such inner states. In other words, following this line of thought, I have no criteria to decide whether others are human or living beings, or whether they are only automata (PI, p. 420).

At first glance, both challenges have some power. Yet, at a closer look, they are not without problems. Let us briefly mention only two of these problematic implications, respectively related to the epistemological and the ontological version of the challenge.

First, as Wittgenstein observes, the epistemological challenge based on pretence entails a grammatical contradiction:

One can say 'He is hiding his feelings'. But that means that it is not a priori they are always hidden. Or: There are two statements contradicting one another: one is that feelings are essentially hidden; the other, that someone is hiding his feelings from me. (LWPP II, p. 35)

---

<sup>2</sup> As Avramides (2001) points out, there is not one sceptical argument, for we should clearly distinguish ancient scepticism, which, despite the generality of doubt, has an approach to knowledge grounded on action, from modern epistemological scepticism. The formulation of the problem of other minds that I consider here is only consistent with the modern framework.



The concept of pretence is grammatically grounded upon the concept of genuine expression: it has its meaning only in contrast to the idea of genuine (or ingenuous) expression, which is therefore presupposed (cf. Ryle 2009). Thus, the concept of generalised pretence, without any contrast with genuineness, is meaningless (cf. Jureshi 2012; Ter Hark 1990).

Secondly, the ontological challenge is based on the assumption of the separation between inner states of mind, which are inaccessible, and outer expressions, which we can perceive. The latter are conceived as signs pointing at the former. Since signs are decoupled from what they designate and since there can be signs indicating at non-existent objects, there can be in principle no certainty as to whether the mental states indicated by outer signs really exist or not, i. e., we cannot know with certainty whether the other is another subject, or rather only an automaton.

Besides being, as Wittgenstein observes, uncanny (PI, p. 420), this inference also entails a contradiction. The argument, indeed, begins with the possibility of pretence, conceived as an intentional activity that only minded beings can accomplish, and eventually arrives at the contradictory conclusion that pretence itself is what challenges the very existence of minded beings. Thus, as Cavell (1999, p. 379) points out, the possibility of pretence seems only to leave room for some kind of epistemological agnosticism, but not for metaphysical scepticism.

## 2 Experiencing and expressing

The understanding of pretence that underlies the different versions of the sceptical challenge concerning other minds seems to be grounded on three main ideas concerning experience and the expression thereof: (1) the assumption that, in order to experience others, I need to have the same access to their experiences as I have to mine; (2) the idea that the relation between behaviour and inner states of mind is a kind of signitive indication, i. e., that expressions point at supposedly independent mental states and that there can be either accordance or discordance between the sign (outer behaviour) and the designated (the inner mental state); and (3) the idea that the privileged access I have to my own experiences is a matter of uniquely certain knowledge. This is an oversimplified view of what experience and expression are, and the revision of such a view is a pre-condition for a proper understanding of pretence.

The more or less explicit idea behind the first assumption is that, in order for there being certainty about other minds (be that epistemological or ontological), I need to have a fully immediate (i. e., not mediated by either language or bodily expressions) access to the other's inner experience. Since it is assumed that I

have such an immediate access to my own experiences, what I would need is to have the same direct access to his/her experiences as I have to mine. As this is not the case, I can never be certain about anything concerning the other's experience. One of the sceptical premises is undeniably correct: I cannot in principle have the same access to the other's experiences as to mine. There is a fundamental asymmetry between how I experience my own pain and how I experience that the other is in pain (Overgaard 2007., pp. 91f.). Indeed, if the opposite was true, the very problem of intersubjectivity and otherness would be meaningless or absurd: if I had the same access to the other's experience as I have to mine, the other would simply be a part of myself (Husserl 1960, p. 109). In other words, experiencing is uniquely first-personal and our experience of others is primarily that of separation (Cavell 1976, p. 260). This is not a merely empirical statement, but rather a necessity for any discourse about *inter*-subjectivity. As Cavell (1976, p. 259) writes, what is stated herewith is the necessity of a *Faktum*, "some terrible or fortunate fact, at once contingent and necessary". The mistake in the sceptical argument, thus, consists in ideally taking the possibility of bypassing such fundamental asymmetry as the only ground to legitimately talk about our experience of others.

Such asymmetry, however, is not rendered by the distinction between direct versus inferential access to experiences. Phenomenological theories argue that, despite such asymmetry, we do have a *sui generis* direct experience of others, and this experience goes under the name of empathy.<sup>3</sup> Introducing the concept of imponderable evidence to address the way we grasp other's expressions (e. g., PI II p. xi, p. 228; LWPP I, p. 936), I believe that Wittgenstein would also agree with the phenomenological remark that I do have a specific direct (i. e., non-inferential) experience of others based on their expressions, and I can *see* the anger, sadness, or joy in and through their faces (cf. Scheler 1973, p. 209f.). Being unproblematic, and even essential in our everyday life, imponderable evidence becomes problematic only in the philosophical discourse, for it does not grant any kind of criterial knowledge.<sup>4</sup> It rather designates the intuitive and morphological grasp that we can have of something like a style, which cannot be fully explicat-

---

<sup>3</sup> Referring to empathy, of course, does not simply solve the problems. The debate on the nature of empathy as direct perception of others is particularly rich and it would be impossible to address it here in detail. What I want to emphasize, however, is that empathy is conceived as a *sui generis* intuitive act of presentification, which is directed to the other and his/her experiences. Cf., among many others, Husserl (1960, 1989), Stein (2008), Zahavi (2014).

<sup>4</sup> Ter Hark (1990, p. 154f.) identifies three main features characterising imponderable evidence: (1) it is not based on clearly definable criteria, (ii) it is unpredictable, (iii) there is no full agreement about it.

ed in objectively fixed terms. Yet, imponderable evidence is the only way in which something constitutively indeterminate (irregular and unpredictable) is given to us, like, precisely, the expression of a face or the characteristic atmosphere of a context or situation (Ter Hark 1990, p. 149 f.).<sup>5</sup>

As far as the second assumption is concerned, Moran (2005, p. 351) correctly points out that the expression of one's inner states of mind can be conceived in at least two ways: first, as a "manifestation of some attitude or state of mind", and secondly, as "the intentional act of one person directed to another".

The former corresponds to what I would call primary expression: our remorse, shame, or gratitude manifest themselves (e.g., through our body or the prosody of our speech) without any explicit intention from our side. Such expressions are not under our control, and do not require us to take any stance with respect to our experiences. They cannot be considered as decoupled from what they express, rather being constitutive moments of the experience itself. The cry is not an intimation of pain, but rather belongs to the complex experience of feeling pain. As Finkelstein (2003, p. 135) points out: "A pain and its expression make sense together in something like the two parts of a single sentence do." The co-belonging of experiencing and primary expression is another point in which the implications of Wittgenstein's remarks come close to the phenomenological understanding of lived experience. Being the bearer of primal expressions, the body cannot be considered to be a kind of physiological machine, as the sceptical argument seems to imply. Moreover, mind and body are not two decoupled entities that somehow come into relation with each other, but rather make up a unity of intertwinement (cf. Husserl 1989, p. 99; Summa 2014, pp. 247 f.). In this sense, Wittgenstein's remark that "[a]n inner process stands in need of outward criteria" (PI, p. 580) may also be considered as related to the intertwined unity of mind and body, and to the co-belong-

---

<sup>5</sup> Different from machines, "with a human being, the assumption is that it is impossible to gain insight into the mechanism. Thus indeterminacy is postulated" (RPP II, § 666). "The question of evidence for what is experienced has to be connected with the certainty or uncertainty of foreseeing someone else's behaviour. But that's not quite the way it is; for only rarely does one predict someone else's reaction" (LWPP II, p. 65) "I think unforeseeability must be an essential property of the mental. Just like the endless multiplicity of expression" (LWPP II, p. 65) "The important fine shades of behaviour are not predictable" (LWPP II, p. 65) "I am for instance convinced that my friend was glad to see me. But now, in philosophizing, I say to myself that it could be otherwise: maybe he was just pretending. But then I immediately say to myself that, even if he himself were to admit this, I wouldn't be at all certain that he isn't mistaken in thinking that he knows himself. Thus there is an indeterminacy in the entire game. One could say: In a game in which the rules are indeterminate, one cannot know who has won and who has lost" (LWPP II, p. 86)

ing of experiencing and primal expression. Inner processes and outer expressions are the two sides of the same coin and can only be considered in their unity. Due to such fundamental co-belonging, my experience is not limited to privateness: through primary expressions, it becomes manifest for others as a response to a particular situation. In this respect, Husserl also claims that expressive movements ground a form of bodily communication which allows us to experience others as living and experiencing beings (cf. Husserl 1989, p. 248f.).

The latter kind of expressions Moran considers are acts through which we give expression to our own experiencing with the intention to intimate something to someone. This requires us to take a stance with respect to what we are experiencing. These expressions shall be understood as specific language games grounded on primary expressions. As such, they are of crucial importance to understand pretence.

This brings us, finally, to the third assumption mentioned above. Is such “taking a stance” a matter of knowledge and certainty about our own mental states? This seems to be the claim of the sceptic, whose argument is based on the idea that I know my own mental states better than anyone else. Despite recognising that I have a privileged access to my own experiences, there are several points that speak against such a view. Let me just mention two. The first point derives from the very sceptical argument. As Descartes points out,<sup>6</sup> the certainty of our experiences only extends as long as the experiences are actually present. Accordingly, such certainty could be called into question as soon as the temporal spreading-out of experience is taken into account. And *a fortiori* we could not rely on the certainty of our memories.<sup>7</sup> The second point is that our primary expressions can reveal to others some aspects of our own experiencing we are not aware of, and which we arrive at recognising only in the interaction with others. One could say that there is some excess of expressions over intentions. This is particularly true as far as our emotional experience is concerned. We can thus agree with Overgaard’s (2007, pp. 16f.) view that there is too much intimacy in one’s own self-experience to be translated into the language of knowing. Our own experiences are not inner things, already univocally defined in themselves, which can be observed and made into the object of knowledge. Rather, their intimacy is affective. It is the intimacy of what Waldenfels (e.g., 2002, 2006) calls a *Widerfahrnis*: something in ourselves that surprises us and that we are called to respond to.

---

<sup>6</sup> See, notably, the *Second* and the *Sixth Meditation*, Descartes (1985).

<sup>7</sup> On the basis of a different account of inner time consciousness and of his understanding of the phenomenological reduction, despite inheriting some of the Cartesian epistemological questions, Husserl understands the certainty and the evidence of self-experience (including memory and time-consciousness) in a different way. I have addressed these issues in *Summa* (2014).

The problem, then, is to determine which meaning the concept of knowledge can still have in this context (Cf. PI, p. 246; PI II p. ix; PI II p. xi, pp. 221 f.; LWPP II, p. 92). According to Cavell (1976, p. 255), when used in relation to experience, the concept of knowledge has a shade of meaning that is not related to the certainty we acquire when some definite criteria are fulfilled, but rather to the *acknowledgement* of the experience we (or someone else) are undergoing.<sup>8</sup> What interests us, when we are confronted with experiences (be that our own or the other's), is not the certainty of criterial knowledge, for we actually use expressions like “I know I am in pain”, “I know I am late”, “I know I am childish” when we want to admit or confess something:

It is obvious enough, but unremarked, that “I know I am in pain” (containing the assertible factor “I am in pain”) is an *expression of pain* (accepting Wittgenstein's view that “I am in pain” is such an expression). It may also be an expression of exasperation. And it has the form (“I know I...”) of an acknowledgement [...]. As an acknowledgement (admission, confession) it is perfectly intelligible. (Cavell 1976, p. 256)

This idea is not meant to simply substitute the epistemological concept of criterial knowledge, but rather to indicate that the concept of knowledge has several inflections, among others those related to certainty and those related to acknowledging. These remarks, thus, do not aim at a refutation of scepticism, which has already proven to entail some contradictory points. Rather, they aim to pave the way for a different understanding of our experience of ourselves and others: according to such an understanding, the epistemological questions are somehow to be re-assessed on the basis of how otherness primarily affects us.

### 3 Pretence and intersubjectivity: Toward a relational understanding

Pretence can be reassessed on the basis of the three points discussed in the previous section.

As far as asymmetry is concerned, there is a sense in which, despite our perception of others, we can be mistaken about what s/he is actually experiencing, and about his/her intentions. Like perception in general, our perception of oth-

---

<sup>8</sup> A comparison of Cavell's theory of acknowledgment and Waldenfels's understanding of the “truth of confession” [*Geständniswahrheit*], based on his idea of responsiveness, would be very fruitful in order to better determine the nature of “knowledge” that is at stake here. Cf. Waldenfels (2007, pp. 398–412)

ers is corrigible. Yet, different from the perception of things, the possibility of being wrong about what others experience is not only due to potential illusions from the perceiver's side, but also to the fact that the other may intentionally deceive us and that there is something in the experience of others that, different from the perception of things, remains in principle inaccessible, although we are aware of it "as inaccessible" (cf. Husserl 1960, p. 114). Thus, if we remain within the framework of certainty, recognising whether the other is expressing his/her genuine state of mind or only pretending can only be a matter of imponderable evidence. As such, this evidence is principally fallible, unpredictable, and not based on fully definable criteria (cf. Jureschi 2012).

Having said this, I believe it is more fruitful to address pretence in relation to the second and the third of the above mentioned points, namely in relation to the different forms of expression, and to the affective and responsive nature of our experience of ourselves and others.

With regard to expression, Wittgenstein considers pretence as a language game that needs to be learned. Accordingly, pretence is not simply to be put on a par with feelings or mental states, i. e., it is not true that we have expressions indicating either genuine or pretend states (LWPP II, p. 59). The simple alternative between expressions of genuine experiences and expressions of pretend experiences is not tenable. Instead, pretence should be considered as a practice or a game: "what goes on in him is also a game, and pretence is not present in him like a feeling, but like a game" (LWPP II, pp. 31–32).

The idea that pretence is a language game that needs to be learned is not simply a developmental-psychological claim, but rather refers to the characteristic complexity of such a phenomenon. Pretence does not belong to the domain of what we have called primal expressions, being rather grounded thereupon (Canfield 2004; Ter Hark 1990). As Wittgenstein points out, "a dog cannot be a hypocrite, but neither can he be sincere" (PI II, p. xi, p. 229, cf. also LWPP, I, p. 862). And: "Feigning and its opposite exist only when there is a complicated play of expressions (Just as false or correct moves exist only in a game)" (LWPP I, p. 946). Thus, the concepts of genuine versus feigned do not apply to the primal expression of feelings or other mental states, but only to the grounded level of intimation. And what is also important is that they require intersubjectivity as the consciousness of an addressee, of what s/he might think about us:

What does a child have to learn before he can pretend? Well, for example, the use of words like: 'He thinks I'm feeling pain, but I'm not'. (LWPP, I, p. 866; see also LWPP, I, p. 869).

The concepts of sincerity and pretence only apply to the level of intimation, and only insofar as intimation entails the acknowledgement (rather than the criterial

knowledge) of what we are experiencing and the acknowledgement of the other as interlocutor (cf. Moran 2005). We are in any case sincere when we give voice to a feeling that we acknowledge as ours, even though we might subsequently discover ourselves being wrong about what we thought we were experiencing, maybe precisely through the interaction with others.

I contend that the language game of pretence has an intersubjective and contextual or situational nature.<sup>9</sup> And the role of intersubjectivity here is not only related to the fact that there is always an addressee in front of whom we dissimulate something. Rather, intersubjectivity shall be understood as grounding the very meaningfulness of pretend actions and expressions.

Pretence is not just an intentional manipulation of behaviour based on the knowledge of an inner state of mind, as if the descriptive function of language would completely take over the role of the expressive function. Rather, the complex language game of pretence is a practice that implies a conscious modification of our basic expressiveness, and presupposes some form of responsiveness and acknowledgement of our experiences in front of others. Moreover, pretence constitutively depends on our grasp of the situation and of others as partners or addressees in this situation, i. e., as having a unique point of view. It also implies an imaginative grasp of different scenarios that would possibly open up, should we act or not in a way that is consistent, for instance, with the feeling we acknowledge. Finally, having made the distinction between primary expression and intimation, one shall also emphasise that primary expression does not cease when intimation begins. On the contrary, due to the excess of expression I mentioned above, elements of primary expression, in their intimate connection with what we are experiencing, also emerge in the context of intentional intimation, and even in pretending. The tone of our voice may betray our embarrassment, even if we are trying to appear as fully self-confident. Even in pretending or playing a role, our own expressiveness considered as a whole is not fully under our control. And even explicitly intimating expressions entail more than what we want to say: there is something in them that is not part of what we intentionally want to communicate and that possibly withdraws from our conscious awareness. There are, in other words, unnoticed and uncontrollable aspects in our expressions, which may even be grasped by others, independently of what our communicative intentions are.

Pretence is a form of articulated responsiveness to something that we primarily experience as affection, and that we acknowledge in an intersubjective

---

<sup>9</sup> Wittgenstein, for instance, remarks that we need to have motives or reasons to lie in a particular situation. LWPP I, p. 314. See also Jacquette (2004).

context. Such responsiveness and the modification of what would be a sincere intimating expression are grounded on a double grasp of the experienced situation. On the one hand, a perceptual grasp of the intersubjective context, together with what we can call the imponderable evidence of an atmosphere, for instance, of trust (or the lack thereof). On the other hand, an imaginative grasp related to possible scenarios and to the way one's expressions may be seen from the other. Accordingly, pretence requires us to assume what Plessner (2003, 2010) calls an "eccentric positionality" with respect to our basic and direct experiencing, i.e., it requires us not to lose ourselves in our primal expressive responses. Pretence, thus, has a relational and imaginative nature. The idea that the meaningfulness of something like pretence (and conversely sincerity/genuineness) is relational is grounded on the adoption of such an eccentric position in an intersubjective context. In taking distance from the centrality of my direct experience, I establish a relation to myself and to others, which allows me to possibly take a stance with respect to such experiencing. Yet, in pretence, such a self-decentering is also relational because it cannot be established apart from a given context in which I am exposed to and interact with others. As a language game, it is a practice based not only on expression, but also on acknowledgement, on the grasp of an intersubjective context, of the way my expressions would fit that context, of how they can be received by the other, etc. And, at the same time, it implies an imaginative grasp of possible alternative contexts.

## 5 Conclusions

My aim in this paper was to investigate the phenomenon of pretence in order to reassess its relation to intersubjectivity. I began by considering how such a relation is generally understood, namely by discussing how pretence can be assumed as the basis to challenge both our knowledge and the very idea of existence of other minds. In both cases, I have shown that the sceptical argument entails some contradictory statements. Even if they also address ontological aspects, the sceptical arguments primarily move within the framework of an epistemological question, related to the extents and the limits of our possible knowledge. One way to respond to such a challenge with respect to other minds would be to argue that such an epistemological question is not at all the crucial one. Rather, we should address the problem of other minds as a conceptual one, namely as related to the extension of mental concepts, as pertinent to both ourselves and the others. According to Avramides (2001), who defends such a view, this is also Wittgenstein's approach to the matter.



Although I would agree that this is an important point for the understanding of Wittgenstein's philosophy, one of the plus-values in bringing Wittgenstein in dialogue with the phenomenological tradition may be that of not simply dismissing the epistemological question, and rather phrase it differently. And I believe that a phenomenologically informed inquiry into pretence, considered in relation with asymmetry, expressiveness, and acknowledgement, allows such a rephrasing, while remaining at the same time faithful to Wittgenstein's positions.

According to such a reformulation, any discussion of how we experience others needs to begin with the fundamental experience of separation, of the alterity of others, and of the constitutive asymmetry between self- and other-experience. Moreover, it has to take seriously both the co-belonging of experience and primal expression and the distinction of the latter from intimation. In this sense, the intuitive grasping of the other through his/her primal expressions is what grounds any form of possible knowledge about the other. In a second step, we may try to make what we grasp explicit, express it as a form of "knowledge", and even try to identify criteria for such knowledge. Yet, since the evidence we have of primal expressions is imponderable, such criteria cannot be exhaustive, and our knowledge cannot be certain.

Ordinary language seems however to reveal that the kind of knowledge we have while expressing (intimating, confessing) something about ourselves or while noticing something in the other's expressions is not primarily a matter of criterial identification, but rather a matter of acknowledgement. Only if I acknowledge my experience for myself and in front of others, and only if I have already acknowledged others as partners of interaction, can I be sincere or rather pretend. And this is at least partially independent of criterial knowledge and its correctness. While being sincere or pretending, the language game we play is one that already entails our basic recognition of the other as a partner. In our ordinary dealing with others, the way we grasp their expressions is a form of acknowledging, or a response to a claim someone makes upon us (Cavell 1976, p. 263). This, I would argue, is based on the imponderable evidence of such claim.

## Bibliography

- Anscombe, Elizabeth (1958). "Pretending". In: *The Aristotelian Society Supplementary Volume XXXII*, pp. 279–294.
- Austin, John Langshaw (1958). "Pretending". In: *The Aristotelian Society Supplementary Volume XXXII*, pp. 261–278.
- Avramides, Anita (2001). *Other Minds*. London, New York: Routledge.
- Canfield, John (2004). "Pretence and the Inner". In: Moyal-Sharrock, Danièle (Ed.): *The Third Wittgenstein: The Post-Investigations Works*. Ashgate: Aldershot, pp. 145–158.

- Cavell, Stanley (1976). *Must We Mean What We Say? A Book of Essays*. Cambridge: Cambridge University Press.
- Cavell, Stanley (1999). *The Claim of Reason. Wittgenstein, Skepticism, Morality, and Tragedy*. Oxford: Oxford University Press.
- Descartes, René (1985). *Meditations on First Philosophy*. Cambridge: Cambridge University Press.
- Finkelstein, David (2003). *Expression and the Inner*. Cambridge MA, London: Harvard University Press.
- Husserl, Edmund (1960). *Cartesian Meditations*. Den Haag: Martinus Nijhoff.
- Husserl, Edmund (1989). *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. Second Book. Studies in the Phenomenology of Constitution*. Dordrecht/Boston/London: Kluwer.
- Jacquette, Dale (2004). "Wittgenstein on Lying as a Language-Game". In: Moyal-Sharrock, Danièle (Ed.): *The Third Wittgenstein: The Post-Investigations Work*. Aldershot: Ashgate, pp. 159–176.
- Jureschi, Livia Andreia (2012). "Wittgenstein, Pretence and Uncertainty". In: Padilla Gálvez, Jesús and Gaffal, Margit (Eds.), *Doubtful Certainties. Language Games, Forms of Life, Relativism*. Heusenstamm: Ontos, pp. 91–108.
- Mercolli, Laura (2012). *So tun, als ob. Analyse und Systematik eines ungewöhnlichen Begriffs mit einer Anwendung auf Theorien der Fiktionalität*. Paderborn: Mentis.
- Moran, Richard (2005). "Problems of Sincerity". In: *Proceedings of the Aristotelian Society* 105 (3): pp. 341–361.
- Overgaard, Søren (2007). *Wittgenstein and Other Minds*. New York: Routledge.
- Plessner, Helmuth (2003). *Ausdruck und menschliche Natur*. Frankfurt/Main: Suhrkamp.
- Plessner, Helmuth (2010). *Die Stufen des Organischen und der Mensch*. Berlin/New York: De Gruyter.
- Ryle, Gilbert (2009). *The Concept of Mind*. London: Routledge.
- Scheler, Max (1973). *Wesen und Formen der Sympathie*. Bern/Munich: Francke.
- Stein, Edith (2008). *Zum Problem der Einfühlung*. Freiburg: Herder.
- Summa, Michela (2014). *Spatio-Temporal Intertwining. Husserl's Transcendental Aesthetic*. Dordrecht/Boston/London: Springer.
- Summa, Michela (2016). "Imaginative Dimensions of Reality: Pretense, Knowledge, and Sociality". In: Reynolds, Jack and Sebold, Richard, *Phenomenology and Science*. New York: Palgrave MacMillan, pp. 205–224
- Ter Hark, Michel (1990). *Beyond the Inner and the Outer. Wittgenstein's Philosophy of Psychology*. Dordrecht/Boston/London: Kluwer.
- Waldenfels, Bernhard (2002). *Bruchlinien der Erfahrung*. Frankfurt/Main: Suhrkamp.
- Waldenfels, Bernhard (2006). *Grundmotive einer Phänomenologie des Fremden*. Frankfurt/Main: Suhrkamp.
- Waldenfels, Bernhard (2007). *Antwortregister*. Frankfurt/Main: Suhrkamp.
- Witherspoon, Edward (2011). "Wittgenstein on Criteria and the Problem of Other Minds". In: Kuusela, Oskari and McGinn, Marie (Eds.), *The Oxford Handbook of Wittgenstein*. New York: Oxford University Press, pp. 472–498.
- Wittgenstein, Ludwig (1980). *Remarks on the Philosophy of Psychology*. Oxford: Blackwell.

- Wittgenstein, Ludwig (1982). *Last Writings on the Philosophy of Psychology*. Volume I. Oxford: Blackwell. [LWPP I]
- Wittgenstein, Ludwig (1986). *Philosophical Investigations*. Oxford: Blackwell. [PI]
- Wittgenstein, Ludwig. (1992). *Last Writings on the Philosophy of Psychology*. Volume II. Oxford: Blackwell. [LWPP II]
- Zahavi, Dan (2014). *Self and Other. Exploring Subjectivity, Empathy, and Shame*. Oxford: Oxford University Press.

Hanne Jacobs

# Socialization, Reflection, and Personhood

**Abstract:** According to a predominant view, reflection is constitutive of personhood. In this paper I first indicate how it might seem that such an account cannot do justice to the socially embedded nature of personhood. I then present a phenomenologically-inspired account of reflection as critical stance taking and show how it accommodates the social embeddedness of persons. In concluding, I outline how this phenomenological account is also not vulnerable to a number of additional challenges that have been raised against accounts that consider reflection to be constitutive of personhood and what matters for responsibility.

## 1 Problems with Reflection

The idea that we constitute ourselves as persons by reflecting has intuitive appeal. For example, it seems plausible that reflecting on two conflicting desires, deliberating about which one to identify with, and then identifying with one desire over another are what allow us to become our own persons in some respect.<sup>1</sup> At the same time, it is factually the case that persons are socially embedded—that is, a person's beliefs, evaluative attitudes, and practical projects are to a significant extent informed and shaped by other persons and the socio-historical context more broadly understood.<sup>2</sup> Localizing personhood in the capacity to reflect might seem to entail that one regards this social embeddedness as something extraneous to the autonomous self and perhaps even as something to be overcome. Moreover, once we recognize that persons are socially embedded, it is not clear why a procedure of reflection would lead us to become our own person in some respect. For, as several have critically pointed out in response to procedural accounts of autonomy through reflection, it is not necessary that one's second-order reflective identification and endorsement be one's own.<sup>3</sup> Re-

---

<sup>1</sup> See, most prominently, Frankfurt 1971.

<sup>2</sup> This is a rather uncontroversial commitment in the phenomenological tradition. See, for example, Carr 2014, p. 48f.; Chelstrom 2013, p. 60f.; Steinbock 1995, p. 186f. The idea is also now well entrenched in analytic discussions on personhood (for an overview, see Christman 2009, chap. 1).

<sup>3</sup> See, for example, Benson 1994 and 2005, Oshana 1998, Stoljar 2000, and Meyers 2000. See also the discussion in Christman 2009, chapter 6.

sponding to this criticism, procedural accounts of autonomy have taken up the challenge of showing how a modified procedural account of reflection can compensate for this shortcoming.<sup>4</sup>

In this paper, I provide a phenomenologically-inspired account of what kind of reflection is constitutive of personhood. My aim is to show that this account can do justice to the social embeddedness of persons in a number of important respects. Further, in concluding I suggest that this account can also stand up to some other challenges that have been raised against the claim that reflection is constitutive of personhood and what matters for responsibility.

## 2 Reflection and Intentionality

The account of reflection that I present in what follows draws on Husserl's phenomenological account of intentionality and personhood. While Husserl usually uses the term "reflection" to refer to a second-order reflection on conscious experience,<sup>5</sup> this is not the kind of reflection that is constitutive of personhood. Rather, as I would like to emphasize, what is constitutive of personhood is the specific kind of reflexivity that takes the shape of what Husserl calls critical stances (*Stellungnahme*).<sup>6</sup>

This reflexivity that characterizes the act of taking a critical stance can be understood in light of the structure of intentionality.<sup>7</sup> On Husserl's account of in-

---

<sup>4</sup> See Christman 2009, pp. 127–63. For a recent overview of number of different positions and their respective difficulties when it comes to the socially embedded nature and autonomy of the self, see Killminster 2013.

<sup>5</sup> See, for example, Husserl 1976, pp. 162–69.

<sup>6</sup> Husserl's use of the term "*Stellungnahme*" is ambiguous (see Drummond 2007, pp. 165–66). In what follows, I use "critical stance" as a translation for *Stellungnahme* in the narrow sense as denoting the critical and deliberative stances that are constitutive of personhood and our standing convictions, appraisals, and volitions. Husserl explicitly characterizes personhood in this way in Husserl 1950, pp. 100–102; Husserl 1962, pp. 206–15; Husserl 2004, pp. 103–24. Other texts that are relevant in this respect are research manuscripts that are part of the forthcoming *Studien zur Struktur des Bewusstseins*, the manuscripts that are collected in the three volumes *Zur Phänomenologie der Intersubjektivität* (Husserl 1973a, nr. 15, nr. 16, Beilage LII; Husserl 1973b, nr. 2, nr. 10; Husserl 1973c, nr. 24, nr. 28, and nr. 29), and the ethical reflections in the recently published *Grenzprobleme der Phänomenologie* (Husserl 2014, nr. 21–22). Individual and dispersed discussions of *Stellungnahme* in the sense of critical stance taking in absence of an explicit discussion of personhood are even more numerous (e.g., Husserl 1939, pp. 325–80; Husserl 1966, pp. 51–64, 357–61; Husserl 1974, pp. 125–35; Husserl 1976, pp. 192, 218–19).

<sup>7</sup> For a more detailed exposition of Husserl's conception of stances, see Jacobs 2010 and Rinofner-Kreidl 2011. Drummond (2010) follows Husserl (1974) by speaking of a critical or proposi-

tentionality, every experience is an experience of something, and this something is always intended in a certain way.<sup>8</sup> That is, the objects of our intentional awareness are always afforded with a certain sense, which is to say that they are always intended *as* something. Objects and situations are not only afforded as having certain natural properties (e. g., as having certain material characteristics); they are also afforded with evaluative and practical properties (e. g., as beautiful, pleasurable, disgusting, useful, or desirable). And objects and situations appear as having certain evaluative or practical properties in virtue of the natural qualities they have.<sup>9</sup> The sense with which objects and situations appear to me is informed by my past experience, my particular embodiment, and the specific context in which I encounter an object or situation, which is, as I discuss in what follows, factually a socio-historical context. Further, when experiencing something with a certain sense, I usually posit or take what I intend in this way to be that way (unless I am merely imagining or contemplating something).<sup>10</sup>

Our ways of intending the world are most often confirmed and refined as we go, which is to say that situations and things usually turn out to be how we take them to be. However, due to the fact that we always intend the world in a way that posits more than what we can currently have the evidence for, our way of intending can be wrong, and there are occasionally situations in which we become aware of the possibility that how we intend something might not be how it is. While the course of experience itself often resolves these moments of doubt,<sup>11</sup> an adult person's experience is regularly characterized by moments of doubt that are not so easily resolved. This is in part because many of the things we believe, appraise, and strive to realize through projects and actions

---

tional attitude, though he takes issue with the idea of deciding or stance taking in terms of affirmation or denial. Rinofner-Kreidl has also proposed her own phenomenological model of self-determination, while also pointing to the different ways in which the autonomous self is socially embedded; on this model, stances are the lowest level of self-determination and are characterized as “eigentlichen und freien Wollen, wenn ich meine Wünsche und Bedürfnisse einer *Prüfung* unterziehe, indem ich mich frage, ob das unmittelbar Erstrebenswerte das ist, was ich in Übereinstimmung mit meinen Wertüberzeugung tatsächlich will bzw. was ich wollen soll” (2012, p. 264, my emphasis). The second level is the level of practical identity and commitment (*Selbstbindung*) and the third level is one of practical self-realization (*Selbstverwirklichung*). In what follows I develop what this “*Prüfung*” or scrutiny might amount to and how it informs our perception even when not so scrutinizing.

**8** For Husserl's account of intentionality, see Drummond 2003.

**9** For an elaboration of Husserl's sophisticated account of the intentionality and rationality of the emotions, see Drummond 2004 and 2013.

**10** See Husserl 1976, pp. 229–34 and 238–42; Husserl 1984, pp. 425–31.

**11** This is what Husserl calls modalization on the level of passivity. See Husserl 1966, p. 33–39. See also Husserl 1939, pp. 93f., 356.

are so complex that even though we could have evidence for why we believe, appraise, or want this rather than that, this evidence is not readily available or conclusive. In such cases, we are compelled to deliberate about what is the case, what is valuable, or what is a worthwhile end.<sup>12</sup> The aim of this deliberation is to take a stance, which amounts to confirming that how we intend something is indeed, after all, or after all not, how we take it to be.<sup>13</sup> Thus, critical stances are a response to a moment of doubt in which we become aware of a possible disagreement between how we intend something and what is intended. I would like to suggest that the awareness of the possibility of disagreement introduces a specific kind of reflexivity in which we become explicitly aware of how we intend something—more precisely, we become aware of the potential disagreement between how we intend something and what is intended. And this disagreement is settled the moment we make up our minds and, so to speak, steady our positing (or reject it) through taking a stance.<sup>14</sup>

The specific intentional modification of reflection in the form of critical stance taking is made possible by the aforementioned essential characteristics of the structure of intentional experience. First, it is only because our intentional awareness is sense-disclosing that the kind of critical reflexivity or scrutiny characteristic of the process of critical stance taking can arise in the first place. That is, because we always intend something with a certain sense, or *as* something, there is always the possibility that how we intend something is not true to what is intended in this way. In a moment of doubt, we become explicitly aware of this possibility. Importantly, the reflexivity introduced by doubt is different from a second-order reflection in that it leaves us geared into the world. When in doubt, I do not reflect on my intentional experience but instead wonder: Is what I intend indeed how I take it to be? Second, it is also due to the structure of our intentional awareness that we cannot but respond in some way or another to such a moment of doubt. Even before reflection, we do not just intend something with a certain sense; we also posit or, better, take what we intend to be as we intend it to be. Doubt inhibits taking something to be a certain way, and if this doubt is not resolved on its own, I cannot but respond to this doubt by either reinforcing this positing, thus validating my way of intending

---

**12** Husserl speaks in this respect of critical stance taking in the strict sense: “Das echte Zuetwas-Stellungnehmen, das Stellungnehmen im prägnanten und eigentlichen Sinn, ist Stellungnehmen zu einer Frage, zu einer Anmutung, Zumutung, für oder gegen <die> ich mich entscheide, und zwar je nachdem <wie> die Aktdomäne ist und die Art der Zumutungen: Urteilsentscheidung, Gefallensentscheidung, Willensentscheidung [...] (Ms. A VI 12/I 94a). See also Husserl 1966, pp. 51–62; Husserl 1976, pp. 243–44.

**13** See Husserl 1966, pp. 53–55.

**14** See Husserl 1966, p. 56; Husserl 1939, p. 356.

something, or rejecting it. Alternately, our situation might require that we hold our positing in abeyance and resolve merely to vouch for the probability of something being such or so.

Following Husserl, I would like to suggest that we understand the capacity for critical stance taking to be constitutive of personhood. That is, I would like to suggest that persons are capable of taking critical stances, and they take stances when they are motivated to do so—specifically, when something speaks against their way of intending something in a certain way.<sup>15</sup> This idea that persons are capable of taking critical stances can be understood as the phenomenological retrieval of the classical idea that persons are capable of reflection and hence rational.<sup>16</sup> According to Husserl, one's way of intending something is rational when how one takes something is true to what is intended in this way.<sup>17</sup> Thus, while we are not just rational in this sense when we take critical stances, I would nevertheless contend that the capacity for stance taking is constitutive of being rational in this sense insofar as it is the other side to being sensitive to evidence for how one intends something even when not taking a stance. That is, if one could in principle not doubt or question how one intends the world to be and then settle this doubt or questioning through the taking of a critical stance, it would not make sense to say that one intends the world in a certain way that can be true to this world and hence rational. This is because my intending the world to be a certain way entails a positing that something is the case, valuable, or worthwhile, and this positing can in principle be disappointed. Doubt is the awareness of the possibility of this disappointment, and the taking of a critical stance is a response to this awareness.

### 3 Socially Embedded Reflection

Having sketched a phenomenologically-inspired account of reflection as stance taking, I would now like to discuss how the capacity for stance taking is constitutive of the autonomy of socially embedded persons. This requires first taking a closer look at the way in which our participation in a larger socio-historical context manifests itself in our straightforward awareness of the world.

As Husserl emphasizes, we, as individual subjects, can participate in a culture because we take over or appropriate (*Übernahme*) ways of being intentionally aware

---

<sup>15</sup> See Husserl 1966, p. 53.

<sup>16</sup> I have discussed how the phenomenological account can be understood in this tradition and how it differs from it in more detail in Jacobs 2016.

<sup>17</sup> See Husserl 1976, p. 314f.



of the world. Our participation in a culture primarily manifests itself in the intentional content of our conscious awareness—that is, in *how* we are intentionally aware of our environment and one another. The myriad ways in which concrete and indeterminate others shape our individual perspectives on something need not imply that a conscious subject could not be intentionally aware of its surroundings in a certain way without belonging to a socio-historical context. Nevertheless, if separated from this context, one's intentional awareness would be strikingly different. So, for example, if we did not belong to a socio-historical context, we would not be aware of particular cultural objects (be they of an artistic or practical nature), we would not perceive the actions of others in the way we do (e. g., as appropriate or inappropriate in a particular situation), and we would not conceive of our futures in terms of specific projects (such as the professions we take up and what it entails to commit to that profession).

In this sense then, persons are always *factually*, even if not necessarily, constituted by the social context in which they find themselves in that how they believe, appraise, and desire is shaped by this context. And we are so constituted because we appropriate ways of intending the world from others. Husserl usually emphasizes that we appropriate through language, both spoken and written, but we also appropriate ways of being intentionally aware from our socio-historical context by taking over bodily practices from others.<sup>18</sup> Everyday practices such as how we use tools and handle objects; how we greet or address one another in certain contexts; and how we express our appreciation of something are all bodily practices that are gleaned from the bodily practices of others.<sup>19</sup> Language just significantly enlarges what we can take over from others beyond the immediate presence of specific others.<sup>20</sup>

Importantly, given the provided account of persons as capable of taking critical stances, the manifold ways in which our experience of the world is informed by others and our social context need not entail that this precludes us from being our own persons in the ways in which we believe, feel, and desire. That is, as I have elaborated in the previous section, we are not only intentionally aware of the world; rather, due to the specific structure of this intentional awareness, we are also sensitive to evidence that speaks for or against taking something in a certain way and, hence, capable of taking stances if so required. And we are our own persons when we do not just intend the world in a certain way

---

**18** See Husserl 1973b, pp. 192–93, 216, 223, and 227; Husserl 1973c, pp. 119, 239, 484–85, 529, and 551; Husserl 2008, p. 344–45; and Husserl 2014, p. 280.

**19** See Husserl 1973b, p. 225; Husserl 1973c, p. 420; Husserl 2008, p. 346.

**20** See Husserl 2008, pp. 347–48.

but have evidence for taking the world in this way.<sup>21</sup> In certain situations this will require actual critical stance taking and the specific kind of reflection and deliberation that it involves. In this way we do not so much come up with our own beliefs, valuations, and projects as we come to understand what it means to believe, appraise, and desire a certain way and acquire insight into what speaks in favor of these ways of intending the world, which might entail that we come to change our minds or challenge entrenched ways of believing, appraising, and wanting. And doing so has an impact on our future experiences of objects and situations and on our actions within a specific context. Negatively stated, emptily and confusedly appropriated beliefs, appraisals, or projects for which I have no evidence will not inform my agency in the world.<sup>22</sup>

Acquiring evidence and taking stances with respect to appropriated ways of perceiving, believing, feeling, and acting is something that someone else cannot do for me. Nevertheless, I believe that the presented account of the reflexivity characteristic of critical stance taking can accommodate how concrete others play at least a twofold role in my taking and upholding stances, which is something that the traditional procedural account of reflection is often considered incapable of doing.

First, the confirmation or rejection by others of what I take to be a certain way in my beliefs, feelings, and projects cannot leave me indifferent because of what it means to believe, appraise, or want something.<sup>23</sup> That is, others take the same world to be a certain way in their beliefs, appraisals, and projects, and, as subjects that experience the same world and situations that I do, they have an equal purchase on what is posited in my believing something to be the case, appraising something as valuable, or considering something to be worth realizing.<sup>24</sup> Now, while my believing, appraising, and wanting something commits me to responding to and justifying myself to others, this does not necessarily entail that I cannot uphold my stances despite disagreement. Nevertheless, as I would like to suggest, our ability to uphold our stances takes for granted another kind of validation from others, which leads to the second sense in which others play an important role in the exercise of our capacity to take stances.

---

**21** Husserl 2014, pp. 266–67.

**22** As Husserl writes: “Wer unweise ist, der wird die Anwendung seiner uneigentlichen Kenntnisse, Fertigungsregeln und Fertigkeiten, die Formen ihrer Übertragung auf neue Fälle leicht verfehlen.” Husserl 2014, p. 268.

**23** This speaks to what Benson (2005) and Westlund (2009) have argued for on different grounds.

**24** Husserl 1950, p. 124; Husserl 1976, p. 102; Husserl 2004, p. 121.

While we cannot but take the world to be a certain way in light of the evidence we have, it is conceivable that not being recognized as reliable (in one or more respects) could have an impact on how we experience the world, especially when it comes to the personal stances we take. Concretely, being pervasively questioned or challenged in a certain respect can conceivably lead one to a form of second-guessing and doubt that undermines what it takes to *uphold* a stance—even if it need not make it impossible for one to uphold a stance or to do so in every respect. So, it is conceivable that certain forms of what has been termed “pernicious socialization” can effectively erode the steadfastness that it takes to uphold stances and to stand up for them.<sup>25</sup>

## 4 Conclusion

In response to a challenge to the view that reflection is constitutive of personhood, I have sketched a phenomenological account of reflection as critical stance taking and begun to show how it accommodates the fact that persons are socially embedded. What I would like to indicate in concluding is that this view could withstand a number of additional criticisms that have been raised against the prevalent conception of reflection as what is constitutive of personhood and what matters for responsibility. First, the prevalent conception of reflection as a kind of second-order identification or endorsement entails that I am only my own person when I have actively identified with a belief, appraisal, or desire (or would do so). Not only does this problematically assume that there is a split in the self insofar as one supposedly identifies with some and not other takings (i. e., the ones one has reflectively endorsed); it also leaves it unclear how the endorsed beliefs, appraisals, or desires have an impact on how one experiences and acts when not reflecting.<sup>26</sup> On the phenomenological account that I have sketched here, however, there is no such problem. For one does not only believe, appraise, and desire for oneself when one reflects and endorses one’s own experiences; rather, one believes, appraises, and desires for oneself when one has evidence for how one takes the world to be. And having evidence for taking the world to be a certain way does not necessarily require reflection in the form of critical stance taking. It is only in those cases where I become

---

<sup>25</sup> This is congenial to Benson’s (1995) point concerning the importance of self-worth. Killminster (2013) introduces the term “pernicious socialization” to denote a form of socialization that undermines autonomy and demands that an account of autonomy explain how pernicious socialization compromises autonomy.

<sup>26</sup> This is Crowell’s critique (2013, p. 256f.) of Korsgaard.

aware that how I intend the world might not be true to how the world is that I am urged to consider what speaks for or against me taking something to be a certain way. Further, settling on a way of intending the world comes to inform my future experience and agency in the world. That is, like any intentional experience, the intentional experience of taking a stance informs my future perceptions by making the sense with which future situations and events disclose themselves to me more clear and distinct, and this comes to inform my agency. And by conceiving of critical reflection as a first-order phenomenon (i. e., not a second-order reflection on an act of consciousness) that can occur in many different respects and on different levels, the proposed phenomenological account makes being a person less demanding and makes room for holding individuals with different capabilities accountable in different respects.<sup>27</sup>

At this point one might wonder whether this phenomenological account of intentionality and critical stance taking does not significantly undercut the importance of our capacity for reflection as critical stance taking. That is, if we are sensitive to evidence that speaks for or against us intending the world in a certain way when not reflecting, and if it is the having of evidence for how one intends the world that makes it that one has not just emptily appropriated a belief, appraisal, or desire, this might seem to imply that stance taking is only tangential, and not constitutive, to being one's own person. However, while it is certainly the case that actual critical stance taking is not constitutive of being one's own person, the *capacity* for taking a critical stance is. That is, as I have already mentioned, it is my view that the capacity for critical stance taking is just the other side to our intending the world in a certain way and our being sensitive to evidence that speaks for or against us taking something in that way. Specifically, the capacity for critical stance taking is entailed by our intending the world in a way that can fall short of what is intended in this way.

A second important concern with the view that reflective endorsement is what is constitutive of personhood is that this conception of personhood has too narrow a view on responsibility. Specifically, according to this view we are responsible for what we have reflectively identified with in an autonomous manner. However, in our everyday practices we do not just hold people accountable for the beliefs, appraisals, and desires that they have endorsed or actively identified with and over which they would thus exercise control.<sup>28</sup> The phenomenological account that I

---

<sup>27</sup> This addresses Benson's (1995, p. 52) worry that higher order theories "underestimate drastically the complexity of the volitional capabilities of many people."

<sup>28</sup> See, for example, P. F. Strawson 1962, Smith 2005, and Hieronimy 2014. One might consider the phenomenological account of stance taking to be an elaboration of what Hieronimy (2014) calls "settling a question."

have sketched here can accommodate these everyday practices, since it can make room for a responsibility for my way of intending the world in the minimal sense that my beliefs, appraisals, and desires are attributable to me even when not reflected on or endorsed by me because I am sensitive to what speaks for or against them independently of actually reflecting. At the same time, however, on the presented account, responsibility remains tied to the capacity for reflection in the sense of critical stance taking. That is, intending the world in a certain way is tied to being sensitive to evidence that speaks for or against intending the world in that way, and being sensitive in this way to what speaks for or against one's intending the world in a certain way implies that one is capable of taking a stance when one is motivated to do so. What is more, on the phenomenological account I have developed, one being responsible in this fundamental way for one's beliefs, emotions, and volitions also makes one responsible for the taking of critical stances in an additional twofold way.<sup>29</sup> Specifically, as subjects capable of taking critical stances, we are responsible for taking such stances when rationally required (i. e., when something speaks against how we take something to be) in addition to being responsible for the actual critical stances we take (i. e., when we deliberate on what is the case, what is valuable, or what is worth pursuing). Thus, when we fail to take a critical stance, we can be blamed for our blindness, gullibility, or insensitivity, and we can be held responsible for our failure to think for ourselves. When, on the other hand, we do think for ourselves but fail to do so in a way that is truthful or appropriate to the situation at hand, we can be held responsible in another sense and can be blamed for taking what is probable as certain, for giving some evidence too much weight, or for failing to take into account other available resources in making up our minds.

Broadening the scope of responsibility beyond what we have reflectively identified with calls for a further discussion of the relation between responsibility, blameworthiness, and excuse, which is a discussion that I do not have the space to engage in here. What I at least hope to have shown is that a phenomenologically-inspired account of reflection does not need to limit our responsibility to what we have endorsed or identified with in an act of reflection. Nevertheless, the capacity for reflection matters for who we are as persons in that those who intend the world in a certain way and are sensitive toward evidence that speaks for or against their taking the world in this way will at times be called on to reflect and take a critical stance—specifically, when there is no immediately conclusive evidence for be-

---

<sup>29</sup> This latter form of responsibility seems to correspond to what Hieronimy (2014) calls “jurisdictional responsibility,” which concern things that you can manage; and we can manage ourselves to be sensitive to evidence as well as sensitive to the degree of warrant of our evidence.

lieving, appraising, or desiring in one way rather than another. And our doing so is nothing but an expression of our rationality and the desire to be true to how things are that is intrinsic to the structure of intentionality. That is, insofar as our experience always already posits the world in a certain way, we are pre-reflectively aware of what speaks for or against our taking the world to be this way. As such, we cannot but respond when what we take to be the case, valuable, or worthwhile seems different than how we took it to be, and we cannot but deliberate about what then is the case, valuable, or worthwhile.

## Bibliography

- Benson, Paul (1994): “Free Agency and Self Worth”. In: *Journal of Philosophy* 91. No. 12, pp. 650–58.
- Benson, Paul (2005): “Taking Ownership: Authority and Voice in Autonomous Agency”. In: John Christman/Joel Anderson (Eds.): *Autonomy and the Challenges to Liberalism: New Essays*. Cambridge: Cambridge University Press, pp. 101–26.
- Carr, David (2014): *Experience and History: Phenomenological Perspectives on the Historical World*. Oxford: Oxford University Press.
- Chelstrom, Eric (2013): *Social Phenomenology: Husserl, Intersubjectivity, and Collective Intentionality*. Plymouth: Lexington Books.
- Christman, John Philip (2009): *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*. Cambridge: Cambridge University Press.
- Crowell, Steven (2013): *Normativity and Phenomenology in Husserl and Heidegger*. Cambridge: Cambridge University Press.
- Drummond, John (2003): “The Structure of Intentionality”. In: Donn Welton (Ed.): *The New Husserl: A Critical Reader*. Bloomington: Indiana University Press, pp. 65–92.
- Drummond, John (2004): “‘Cognitive Impenetrability’ and the Complex Intentionality of the Emotions”. In: *Journal of Consciousness Studies* 11. No. 10–11, pp. 109–26.
- Drummond, John (2007): *Historical Dictionary of Husserl’s Philosophy*. Lanham, MD: Scarecrow Press.
- Drummond, John (2010): “Self-Responsibility and Eudaimonia”. In: Carlo Ierna/Hanne Jacobs/Filip Mattens (Eds.) *Philosophy, Phenomenology, Sciences: Essays in Commemoration of Edmund Husserl*. Dordrecht: Springer, pp. 411–30.
- Drummond, John (2014): “The Intentional Structure of Emotions”: In: *Logical Analysis and the History of Philosophy* 16, pp. 244–63.
- Frankfurt, Harry (1971): “Freedom of the Will and the Concept of a Person”. In: *Journal of Philosophy* 68, pp. 5–20.
- Hieronymi, Pamela (2014): “Reflection and Responsibility”. In: *Philosophy and Public Affairs* 42. No. 1, pp. 3–41.
- Husserl, Edmund (1939): *Erfahrung und Urteil. Untersuchungen zur Genealogie der Logik*. Ludwig Landgrebe (Ed.). Prag: Academia/Verlagsbuchhandlung.
- Husserl, Edmund (1950): *Cartesianische Meditationen und Pariser Vorträge*. S. Strasser (Ed.). The Hague: Martinus Nijhoff.

- Husserl, Edmund (1962): *Phänomenologische Psychologie. Vorlesungen Sommersemester 1925*. Walter Biemel (Ed.). The Hague: Martinus Nijhoff.
- Husserl, Edmund (1966): *Analysen zur passiven Synthesis. Aus Vorlesungs- und Forschungsmanuskripten, 1918–1926*. Margot Fleischer (Ed.). The Hague: Martinus Nijhoff.
- Husserl, Edmund (1973a): *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Erster Teil. 1905–1920*. Iso Kern (Ed.). The Hague, Martinus Nijhoff.
- Husserl, Edmund (1973b): *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Zweiter Teil. 1921–28*. Iso Kern (Ed.). The Hague, Martinus Nijhoff.
- Husserl, Edmund (1973c): *Zur Phänomenologie der Intersubjektivität. Texte aus dem Nachlass. Dritter Teil. 1929–35*. Iso Kern (Ed.). The Hague, Martinus Nijhoff.
- Husserl, Edmund (1974): *Formale und transzendente Logik. Versuch einer Kritik der logischen Vernunft*. Paul Janssen (Ed.). The Hague: Martinus Nijhoff.
- Husserl, Edmund (1976): *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Erstes Buch: Allgemeine Einführung in die reine Phänomenologie*. Karl Schuhmann (Ed.). The Hague: Martinus Nijhoff.
- Husserl, Edmund (1984): *Logische Untersuchungen. Zweiter Teil*. Ursula Panzer (Ed.). The Hague: Martinus Nijhoff.
- Husserl, Edmund (1988): *Vorlesungen über Ethik und Wertlehre. 1908–1914*. The Hague: Kluwer.
- Husserl, Edmund (2004): *Einleitung in die Ethik. Vorlesungen Sommersemester 1920 und 1924*. Henning Peucker (Ed.). Dordrecht: Kluwer.
- Husserl, Edmund (2008): *Die Lebenswelt. Auslegungen der vorgegebenen Welt und ihrer Konstitution. Texte aus dem Nachlass (1916–1937)*. Rochus Sowa (Ed.). Dordrecht: Springer.
- Husserl, Edmund (2014): *Grenzprobleme der Phänomenologie. Analysen des Unbewusstseins und der Instinkte. Metaphysik. Späte Ethik (Texte aus dem Nachlass 1908–1937)*. Rochus Sowa/Thomas Vongehr (Eds.). Dordrecht: Springer.
- Jacobs, Hanne (2010): “Towards a Phenomenological Account of Personal Identity”. In: Carlo Ierna/Hanne Jacobs/Filip Mattens (Eds.) *Philosophy, Phenomenology, Sciences: Essays in Commemoration of Edmund Husserl*. Dordrecht: Springer, pp. 333–61.
- Jacobs, Hanne (2016): “Husserl on Reason, Reflection, and Attention”. In: *Research in Phenomenology* 46/2, pp. 257–76.
- Killminster, Suzy (2013): “Autonomy and the Problem of Socialisation”. In: *Social Theory and Practice* 39/1, pp. 95–119.
- Meyers, Diana Tietjens (2000): “Intersectional Identity and the Authentic Self? Opposites Attract”. In: Catriona Mackenzie/Natalie Stoljar (Eds.): *Relational Autonomy*. Oxford: Oxford University Press, pp. 151–80.
- Oshana, Marina A. L. (1998): “Personal Autonomy and Society”. In: *Journal of Social Philosophy* 29. No. 1, pp. 81–102.
- Rinofner-Kreidl, Sonja (2011): “Motive, Gründe und Entscheidungen in Husserls intentionaler Handlungstheorie”. In: Verena Mayer/Christopher Erhard/Marisa Scherini (Eds.): *Die Aktualität Husserls*. Freiburg im Breisgau: Karl Alber, pp. 232–77.
- Rinofner-Kreidl, Sonja (2012): “Selbstbestimmung: ein Stufenmodell zur Klärung eines fundamentalen Begriffs”. In: R. Esterbauer/M. Ross (Ed.): *Den Menschen im Blick. Phänomenologische Zugänge*. Würzburg: Köningshausen & Neumann, pp. 259–79.

- Smith, Angela M. (2005): "Responsibility for Attitudes: Activity and Passivity in Mental Life". In: *Ethics* 115. No. 2, pp. 236–71.
- Strawson, P. F. (1962): "Freedom and Resentment". In: *Proceedings of the British Academy* 48, pp. 1–25.
- Steinbock, Anthony (1995): *Home and Beyond: Generative Phenomenology After Husserl*. Evanston: Northwestern University Press.
- Stoljar, Natalie (2000): "Autonomy and the Feminist Intuition". In Catriona Mackenzie/Natalie Stoljar (Eds.): *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. Oxford: Oxford University Press, pp. 94–111.
- Westlund, Andrea C. (2009): "Rethinking Relational Autonomy". In: *Hypatia* 24. No. 4, pp. 26–49.





Robert Hugo Ziegler

# ‘Vaulting Ambition’ – Machiavelli’s Empty and Impure Concepts

**Abstract:** In the face of political instability and continuous foreign intervention in Italy, Machiavelli sought new ways to answer some of the central questions of political theory: How can a state acquire stability? How can the political be scientifically grasped? What practical advice can be deduced from this scientific knowledge? In his inquiry, Machiavelli fully accepts the radical historicity of values as well as political systems. This “relativist” approach distinguishes Machiavelli from previous theories, ensures its lasting relevance, and constitutes the biggest obstacle to answering the aforementioned questions.

I will try to show in which way Machiavelli succeeds to overcome this obstacle, how he conceives of the genesis of laws and ethical ideas, and how “impure concepts” contribute to the possibility of social life. My interest in Machiavelli is not primarily historical but systematic and I hope to demonstrate to which extent Machiavelli’s theory of the political can still be of use when we try to understand the relations of fact and value, laws and ethical ideas, and the possibility of liberty. Moreover, Machiavelli’s use of “impure concepts” challenges fundamental methodological presuppositions of both phenomenology and analytical philosophy.

## 1 Vaulting ambition

Empires crumble, societies decline, states vanish, and with them their morality, their virtues, and their beliefs. This truth, as cruel as it is simple, has always been known. But philosophy had its ways of ignoring or at least diminishing the importance of the fundamental contingency of all political and moral order. What is the birth and death of a state in the face of the everlasting Ideas or in the eyes of God? Are the dominant beliefs, cherished by a society, more than just the outward and changeable appearance of a moral order which concerns primarily not the state but the individuals? Are all societies not in the end just the attempts of man to approach the ideal and eternal order of society? Man, in his utter inadequacy, might not be able to reach this ideal order, still it lies before his eyes, for instance in those societies of bees in which every individual is just element of a perfect harmony.

Simple devaluation of the empirical in comparison to supposed transcendences, localisation of the essence of morality in the individual rather than in the po-

litical, interpretation of moral concepts and realities in the sense of approximations to the ideal morality – those are some of the strategies used to take the edge off life in this world, this world of mortal and suffering beings, of transient political orders and changing values. In all of them, we can perceive the tendency (and the wish) to presuppose the Ideal and the Eternal, be it the Nature of man, the Truths of right and wrong or the Harmony of a perfect community.

Niccolò Machiavelli, however, could not content himself with these answers. Lacking belief in those theories that offer comfort by claiming that, even if we cannot always see it, yet there certainly is a sense in history and a providence governing life; witnessing Italy tearing herself apart and, in consequence, being devoured by foreign forces, *Machiavelli perceived the radical contingency and historicity of political and moral order* with an acuteness that caused him deep sorrow, as clearly emerges from his writing.

Machiavelli was probably the first in Western philosophy to make this contingency and historicity the starting point of his theoretical efforts.<sup>1</sup> They were his horror, the motivation of his writing, and the object of his theory. His aim was, quite traditionally, to find ways to cope with this contingency by uncovering its immanent logic and laws. His aim may thus join the tendency I have noted earlier, namely to weaken the empirical and temporal by finding the eternal and essential. But Machiavelli remains one of the most radical political philosophers for at least two reasons: On the one hand, Machiavelli never manages to put aside and neutralize contingency and historicity; how could he: they are his primary objects. His *explicit* effort to come to terms with them attributes a new theoretical dignity to them. On the other hand, Machiavelli is not looking to transcend the given problems into a higher order of being; the laws he will establish are decidedly *immanent* laws, laws and structures of and in *this* world.

I will, in what follows, retrace some of the central ideas of Machiavelli. We will see how, according to Machiavelli, political freedom is possible (2); how the idea of justice makes its appearance in the world (3); and how the lasting relevance of Machiavelli's theories lies in the importance he attributes to impure and empty concepts (4). But first we have to ask *why* this world is subject to all sorts of unrest, rebellion, and instability. For what *reason* is it impossible to establish a peaceful and harmonious society?

The answer to this question brings us to one of the central concepts in Machiavelli's theory: *ambition* (*ambizione*). Even if it were possible to find the per-

---

<sup>1</sup> It is, of course, as the culmination of a long development that Machiavelli can formulate his radical attitude in this respect. For the elaboration of the concept of history from the late Middle Ages on, cf. the first part of: Pocock 1975. For the direct precursors and contemporaries of Machiavelli see: Gilbert 1949, and Gilbert 1965.

fect balance of all interests and claims, soon someone would tip the balance by *wanting more*. Machiavelli mostly talks about ambition in the context of power or riches, but we can define ambition in a completely formal way: *Ambition is the tendency, natural to man, to want more, to go beyond... (all boundaries, limitations...), never to be satisfied*. It is this *dynamic nature of man* that is the root of history. Machiavelli describes the mechanism of ambition in these terms:

For, whenever men do not need to struggle with each other out of necessity, they struggle out of ambition; ambition is so powerful in the hearts of men that, no matter how high they rise, it never leaves them. The reason for this is that nature has created men in such way that they can desire everything, but not attain everything; consequently, their desire will always be greater than their power to attain, from which follows discontent with what they have and only little satisfaction of their desire.<sup>2</sup> (Machiavelli 1997, p. 139–40)

Ambition, in this sense, is not simply the aspiration to power and glory; it is a desire that can attach itself to any given object and that can replace every particular motive. To be precise: Because we *are* our own ambition, we can have more particular motives for our actions. In some liminal situations, this radical structure of motivation may show. This is essentially what Macbeth acknowledges in the verses from which I have taken the title of this paper.

I have no spur  
To prick the sides of my intent, but only  
Vaulting ambition which o'erleaps itself  
And falls on th'other – (I, 7, 25–8)  
(Shakespeare 2008, p. 149)

In the absence of every legitimate motive, ambition is all that remains. When Macbeth realizes that, he is willing to restrain this last and absolutely self-sufficient drive. But just then, Lady Macbeth enters and interrupts his monologue – and Macbeth has to learn that ambition is *contagious*.

When we consider what we heard until now, when we consider in particular the fact that Shakespeare's *Macbeth* could serve as an illustration of Machiavelli's concept of ambition (the affinities go further, but I will leave it at that), we are tempted to think that ambition not only is the root of history, but also the root of

---

<sup>2</sup> Discorsi I, 37: "Perché qualunque volta è tolto agli uomini il combattere per necessità, combattono per ambizione; la quale è tanto potente ne' petti umani che mai, a qualunque grado si salgano, gli abbandona. La cagione è perché la natura ha creati gli uomini in modo che possono desiderare ogni cosa e non possono conseguire ogni cosa; talché, essendo sempre maggiore il desiderio che la potenza dello acquistare, ne risulta la mala contentezza di quello che si possiede, e la poca sodisfazione d'esso." All translations from Machiavelli are my own.

all evil. In fact, Machiavelli's poem *Capitolo dell'Ambizione* paints a grim picture of a world prey to ambition (Machiavelli 1981, p. 142–153). But one wonders: If ambition – in the formal sense we have defined – is the nature of man, is it still possible or does it make sense to bewail this nature and its consequences? Should we not rather state the fact and simply analyse it?

In fact, in his theoretical texts, Machiavelli does just that. However, he does share the fundamental belief of conservative thought: that man has a nature that will prove dangerous as soon as it is left to its own. Therefore, man has to be *disciplined*. Machiavelli envisages a series of techniques to this end, the most important of which are certainly military exercise and, above all, the *laws*. But let us turn to another mechanism that allows for a special kind of “discipline”.

## 2 Conflict and freedom

At no point does Machiavelli define political freedom. At the same time, freedom is doubtless the core value of his political theory. The explanation of this seeming contradiction leads us to one of the most modern and still relevant parts of Machiavelli's theory. We have already seen how he refuses to start from any kind of ideal community in relation to which one could judge all actual communities with regard to their respective quality. His only criterion for the “quality” of a state is its longevity – a purely quantitative criterion.

Refusing all idealistic or utopian approaches, Machiavelli can simply state what every observation of actual societies reveals: society, as we know it, is essentially *conflict* (*disunione*). The conflictual nature of organised societies is not an accident, it is not a simple deficiency, owed to our intellectual limitation or moral inadequacy; conflict is in itself the very essence of societies.

Machiavelli further characterizes this conflict by roughly distinguishing two classes in every society: the people and the great (*grandi*), those who are governed and those in power. These classes are defined in terms of participation in political decision making, in terms of economic potency, but also in terms of a shared ambition: The ambition of the great is to expand their power over those governed, while the ambition of the people is to impede this expansion. We have noted earlier that ambition is contagious; we see now that this contagious nature of ambition is responsible for a certain, at least rudimentary solidarity inside the classes, and that means: for the very constitution of those classes as classes.

And what about freedom? Machiavelli is quite clear about it: Political freedom is something that emerges when the conflict essential to a society is shaped in a way that the respective ambitions of the conflicting classes propel one another, by cancelling out each other (at least partially) and by creating new situa-

tions and relations of power. These situations and relations of power remain precarious, they cannot – as long as freedom is an issue – be brought into a definitive harmony, they change and they carry the weight of their contingent genesis with them, they may be compromised or even break down, or they can lift the society to a peak as yet unheard of. In any case, *freedom is in itself the dynamic effect of the essential conflict of a society*. This is why freedom cannot be defined; it appears as though the very fact that 'freedom' must remain an empty concept is the reason for the relevance and effectiveness of this concept!

Freedom becomes possible when the respective ambitions of the classes discipline one another. In order to allow this mechanism to work, the constitution of the society (however informal) must accept and welcome the conflict, including its more violent and at times even rebellious forms. Machiavelli talks explicitly about the necessity to "give room (*dare luogo*) to the commotions and general quarrels in the best possible way" (Machiavelli 1997, p. 77).<sup>3</sup> And he writes:

I maintain that those who condemn the struggles between the nobility and the people, condemn at the same time the very causes that led primarily to the establishment and the conservation of freedom in Rome. He who pays more attention to the noise and clamour of these conflicts than to their good effects, does not take into account that, in every republic, the people and the great have very different aspirations and that all laws in favour of freedom owe their making to this discord.<sup>4</sup> (Machiavelli 1997, p. 71)

In order for freedom to be possible, a society must brace itself for its own constantly renewed conflict and prepare ways of dealing with the conflict. Otherwise, the conflict might become the downfall of the whole society. In Machiavelli's view, the decisive step in Roman history is the establishment of the office of the tribune of the people (*tribunus plebis*).

In the course of the development of this theory of political freedom, something interesting happens to Machiavelli: While paying tribute to the inherent conflict of human communities, a conflict of another kind enters his theory, and I am not sure if Machiavelli himself realized it: It is nearly a commonplace

---

3 Discorsi. I, 6: "[...] è necessario [...] dare luogo a' tumulti e alle dissensioni universali il meglio che si può [...]."

4 Discorsi. I, 4: "Io dico che coloro che danno i tumulti intra i Nobili e la Plebe mi pare che biasimino quelle cose che furono prima causa del tenere libera Roma, e considerino più a' romori e alle grida che di tali tumulti nascevano, che a' buoni effetti che quelli partorivano; e che e' non considerino come e' sono in ogni republica due umori diversi, quello del popolo e quelle de' grandi; e come tutte le leggi che si fanno in favore della libertà, nascono dalla disunione loro [...]. I translated Machiavelli's "umori" with "aspirations". In fact, *umori* in Machiavelli unites the ideas of opinions, emotions and interests; I don't think there is an adequate translation.

in Machiavelli research, and a well-founded one, that the only aim of any state can be its own conservation. No transcendent end can be invoked when discussing the “quality” of a given state. Nonetheless, while we follow Machiavelli’s discussion of the free republic, we suddenly witness the emergence of another idea of the state, the idea of a qualitatively different and “better” state: Machiavelli explicitly says that it is possible to neutralize the constitutional conflict of societies; his examples are Sparta and Venice. And in fact, he regularly invokes the 800 years of unchanged rule in Sparta. Yet, his main subject matter is neither Sparta nor Venice, it is not the “perfect” state; it is this imperfect state that was Rome in its Republican era. The empty but operative concept of freedom puts forward another idea of the state, trumping the seemingly perfect states like Sparta or Venice, whose longevity bears witness to their “perfection”.<sup>5</sup> Only the imperfect but *perfectible* state can create situations in which political freedom can become reality. And this reality is in itself not a perfection or an ideal, but simply the actual and always precarious ‘compromise’.<sup>6</sup> By creating these situations of freedom, the perfectible state puts itself at risk: The conflict can lead to the ruin of the republic. But at least, freedom has been tried.

### 3 Emotions, laws and justice

So, we have empty concepts, imperfect states and only situational freedom. It gets worse. In chapter 2 of book I of his *Discourses*, Machiavelli tells a story. It is one of those stories trying to explain how people first established organised societies. According to Machiavelli, people at first lived in very small populations and thus far apart from one another (in families or small tribes). When population began to grow, so did the number of possible dangers. Accordingly, the families or tribes joined in larger communities in order to better defend themselves. As mutual protection was the first goal of these primitive societies, their members elected the bravest and most valiant individual as their leader. But then

---

5 Therefore, Gilbert is mistaken when he writes: “[...] he [Machiavelli] was impressed that the Romans had been able to stave off their inescapable fate for centuries – longer than had any other political society” (Gilbert 1965, p. 185). For Machiavelli, the constitutional difference between the Roman Republic and the Empire is such that they cannot be viewed as two successive forms of *one* political society; Machiavelli is almost exclusively interested in the former.

6 The word can be misleading. Once again: It is not about a compromise of interests and claims, but about a newly established relation of power(s) (from which the articulation of “interests” may follow). See for a similar theory Laclau/Mouffe 2014, p. xii: “Indeed, politico-hegemonic articulations retroactively create the interests they claim to represent.”

something changed: In these early societies, it might have happened that one who had helped another member of the society did not receive gratitude from the latter, but instead was harmed by him. As the other members witnessed this *ingratitude*, they felt hatred and compassion, and it got them thinking that the same ingratitude could concern them too one day. In order to prevent this, laws were passed and punishments in the case of infraction were fixed. Machiavelli comments: "from there came the cognition of justice" (Machiavelli 1997, p. 66).<sup>7</sup> The next time the citizens had to elect a leader, they would turn, not to the strongest and bravest, but to the wisest and most just.

It is easy to see that the story Machiavelli tells us is his version of the Biblical fall. How did the knowledge of good and evil (or, as Machiavelli puts it: honourable and good versus pernicious and bad) come about? Let's retrace his steps: At the beginning, there is an act of ingratitude, and it has to be noted that this ingratitude does not yet have the features of true morality; it establishes a kind of proto-morality. As a reaction, the members of the society are subjected to an emotion, and as often with emotions, the emotion in question is blurred, so to speak, or indistinct.<sup>8</sup> In an indistinctness essential to this emotion, it mixes compassion with hatred: The citizens feel two emotions at the same time. But the emotion lacks distinctness in another respect too: The citizens not only feel for the other, they fear for themselves. We can resume: Neither the exact phenomenological 'content' nor the real subject of the emotion are differentiated. This indistinctness is not a deficiency; it is quite simply the manner of being of this emotion.

But the emotion does not stay there. It tends to its own objectification in the form of laws. And now, from these laws, the concept of justice emerges. Machiavelli, in a daring manoeuvre, turns the classical conceptions upside down: The laws do not imitate an eternal order of justice; the very idea of justice is nothing than an *effect* of the establishment of laws. This involves a series of consequences: Justice must be considered as a reflection from positive laws. (And I take "reflection" in the broad sense, uniting the optical, the cognitive and the aesthetic sense.) In no way can justice claim any transcendence; it is not absolute, but *conditioned*. And it is conditioned precisely by contingent events in a given society. Furthermore,

---

7 (Discorsi. I, 2: "donde venne la cognizione della giustizia." I retain the term "cognition" because it is the direct translation of the Italian text.

8 The term better suited here would be "unclear". Lack of clearness means, according to the use inherited from Descartes: We do not know what the exact content of our idea is, what it is "about". On the other hand, lack of distinctness means: We are not (yet) able to perfectly differentiate all important *internal* features of our idea, what it "contains". The more adequate term would thus be "unclear", but the English word is too common to convey the specific meaning. Hence my use of the more metaphorical term "blurred".



the wish to define justice betrays a fundamental misunderstanding concerning its ontological status: Justice is in itself not determined, because it is essentially a reflection of something positive and determined. The concept of justice is as blurred as the emotion, that started this series, but for other reasons. Nonetheless, the two can communicate, as the phenomenon of indignation so clearly proves: Indignation is the short-cut between an emotion and a presumed concept of justice, indicating negatively a real or supposed lack in positive order.

And there is more: Machiavelli holds that we must accept the fact that the realization of justice in societies can never be perfect, that the mere expectation of such a perfect justice, the very idea of perfect justice is the first error of political theory. For he returns to the problem of ingratitude some chapters later, but this time he stresses the positive value of ingratitude: In a healthy republic, suspicion against those who, by their own merits, have reached a well-deserved high position might be unjust and ungrateful; it can nevertheless be extremely important for the conservation of freedom (cf. Machiavelli 1997, book I, chapters 28–30). The occasional ingratitude can thus serve as an effective mechanism in a free republic!

If we follow Machiavelli, we can extract this lesson: We must detach ourselves from the ideality and purity of justice. Justice in this world has to be compromised, in order to be at all. As a concept, justice is blurred (because its ‘content’ cannot be distinctly identified) and impure (because it cannot shake its origins in a contingent history); as principle of law-making, justice must consequently accept its own conditioned and limited status.

## 4 Machiavelli’s contributions to a contemporary political philosophy

We can now summarize the results. Machiavelli’s modernity is due to his particular courage: What he feared most was the ruin of social order. What he experienced the most vividly was the fact that no given political system, not even the most promising, can claim the universal truth of its values and guarantee its own success or even survival. But instead of turning his back to these problems by addressing himself to a supposed ideal order of things, he kept his eyes fixed on them in the hope of extracting some kind of *immanent laws* that govern a world otherwise prey to fortune and chance.

In the course of this project, he develops a theory of political freedom which refrains from all speculation about the essence or substance of freedom. Instead, freedom becomes the effect and result, always temporary, of certain situations in which the fundamental conflict of society is fought out. Those situations may put

the society at risk, but they steer the state away from the much too high demand of perfection by putting its perfectibility to the test. Freedom as a concept has to retain its lack of determination; it is essentially an empty concept.<sup>9</sup>

Likewise, the idea of justice not only lacks any clear determination but must be put back where it belongs: in the series of emotion, laws and justice. The idea of justice is thus the reflection of positive and consequently contingent laws of this or that society. It communicates moreover with an emotion that remains blurred as well. Unable to shake its origins in positive and occasioned laws, in its manner of being a lot nearer to a blurred emotion than to the sharply defined laws, justice must be understood not only as an empty, but as an essentially impure concept. The effort of realizing perfect justice is not only an undertaking that leads much too often to the most violent inhumanity; it is simply a misunderstanding concerning the ontological status of justice.

So, the “value”, if we want to call it that, the “value” of justice depends on the facts of law-making. But maybe we ought to turn around the question of facts and values: Why don’t we just accept justice, freedom and other similar concepts or ideas for what they are: *facts in their very own right!* The empty concept of freedom, the situational freedom of societies in conflict, the blurred vision of justice – they all are quite real, because they actively participate in the shaping of societies. They react upon the situations in which they are originally produced and propel them accordingly. Every fact is contingent; but that is not the interesting part. The question is: Which facts – economic, political, constitutional, but also conceptual or “ideal” facts – can *act* and *in what way* do they act?<sup>10</sup>

In a perspective like the one Machiavelli adopts, reality means exclusively action or the power to act (what the German calls “Wirklichkeit”). It’s his decidedly *immanent view of metaphysics* which imposes this idea on him (an idea that he consequently shares with Spinoza, for instance). Reality is a *dynamic* – for better or worse. And the principle of this dynamic is: there are no a priori limits to the power to act. This is the metaphysical meaning of Machiavelli’s *ambizione*. We

---

**9** In our days, Jacques Rancière suggested a very similar theory of freedom and of what he calls politics (Rancière 1995). An important difference, however, between Machiavelli on the one hand and modern writers such as Rancière or Laclau/Mouffe on the other is the relative fixity Machiavelli attributes to social groups and “identities”.

**10** Compare for a very similar idea Castoriadis 1978, pp. 411–2: “[...] la valeur (même ‘économique’), l’égalité, la justice ne sont pas des ‘concepts’ que l’on pourrait fonder, construire (ou même détruire, comme veut parfois le faire Marx pour la justice) dans et par la théorie. Ce sont des idées/significations politiques concernant l’institution de la société telle qu’elle pourrait être et que nous voudrions qu’elle soit – institution qui n’est pas ancrée dans un ordre naturel, logique ou transcendant. Les hommes ne naissent ni libres, ni non-libres, ni égaux, ni non-égaux. *Nous les voulons* (nous nous voulons) libres et égaux dans une société juste et autonome [...]”

can ask whether this principle convinces us; we can ask whether we feel the need for an anthropological foundation of our political theories; but we also have to ask if *ambizione* is really an *anthropological* concept. For we can see that *ambizione* as a principle is at work not only in “man”, but in processes that transcend the individual and in processes that escape the conscious activity of “man”: Is it not *ambizione* too, the aspiration to exceed one’s own limits, when an emotion tends to transcend itself into a law, and when this law tends to transcend itself into the idea of justice? In this view, emotion, law and justice present the three states of the political, in very much the same sense in which we speak of liquids, solid bodies and gases as of the three states of matter. I suspect that this analogy could take us a long way. And is it not *ambizione* too when Machiavelli professes that the only aim of a state is its own conservation but then transcends this very conception towards a dynamic idea of the perfectible state?<sup>11</sup>

In any event, the value of Machiavelli’s theory in the light of contemporary philosophy is obvious: Both analytical philosophy and phenomenology attempt to start from and to shape unambiguous concepts with their clearly defined meaning. (The reference to the so-called ordinary language does not so much overcome the difficulties in this endeavour as it circumvents or ignores them.) Machiavelli, on the other hand, is not interested in the *meaning* of certain concepts, but in the way they *work*. As a result, we are invited to investigate the ontological nature of certain concepts: impure concepts like “justice”, reflections of a positive and defined meaning that claim universal truth by dissociating their validity from its genesis – a dissociation which, if taken literally, will only leave a pale fiction behind or else will lead to a most dangerous moral extremism; empty concepts like “freedom” that create a space where a shapeless emotion can organise and condense itself, the emptiness of the concept being the very condition of possibility of a movement that will bring forth something new. Moreover, Machiavelli asks the same question of “how it works” with respect to the emotion too: The emotion, blurred by its nature, awaits its own articulation into something objective (laws, for instance). It aspires to something without knowing what it is aspiring to. But the reason for this “ignorance” is not the fact that the emotion, being what it is, presents the irrational part of man, the part real and responsible politics should exclude. Rather, the emotion cannot know where to it tends because its ‘aim’ is of another ontological dignity than itself and is therefore not yet included or even implicated in it. Its articulation has yet to be found or invented.

---

11 This comprehensive sense of a dynamic principle of politics, a “desire of power”, is at the centre of Jean-Luc Nancy’s discussion of democracy (Nancy 2012).

No wonder that the realization of what some insist on calling “ideals” proves impossible: not because one is always forced to “compromise” (and thus compromise the “ideals”) or because some sad reality is bound to come in the way. It’s the very nature of the so-called ideals and ideal values as impure or empty concepts that prevents any realization, for the simple reason that they are already very real: they *act*, and they act in the same world where we are used to see relations of power, exploitation, or greed at work. What they can do is contribute to a new formulation and disposition of problems and relations; the idea of a perfect and pure justice, freedom, or equality, on the other hand, is a contradiction in terms. Thus, Machiavelli prepares a dynamic and immanent philosophy of politics that is yet to be developed.

## Bibliography

- Castoriadis, Cornelius (1978): “Valeur, égalité, justice, politique. De Marx à Aristote et d’Aristote à nous”. In: *Les carrefours du labyrinthe* 1. Paris: Seuil, pp. 325–413.
- Gilbert, Felix (1949): “Bernardo Rucellai and the Orti Oricellari. A Study on the Origin of Modern Political Thought”. In: *Journal of the Warburg and Courtauld Institutes* 12, pp. 101–131.
- Gilbert, Felix (1965): *Machiavelli and Guicciardini. Politics and History in Sixteenth-Century Florence*. Princeton: Princeton UP.
- Laclau, Ernesto/Mouffe, Chantal (2014): *Hegemony and Socialist Strategy. Towards a Radical Democratic Politics*. Second Edition. London, New York: Verso.
- Machiavelli, Niccolò (1981): *Capitoli*. Introduzione, testo critico e commentario di Giorgio Inglese. Rome: Bulzoni.
- Machiavelli, Niccolò (1997): *Discorsi sopra la prima deca di Tito Livio*. Introduzione di Gennaro Sasso. Premessa al testo e note di Giorgio Inglese. Milan: RCS (Rizzoli).
- Nancy, Jean-Luc (2012): “Begrenzte und unendliche Demokratie”. In: *Demokratie? Eine Debatte. Mit Beiträgen von G. Agamben, A. Badiou, D. Bensaid, W. Brown, J.-L. Nancy, J. Rancière, K. Ross und S. Žižek*. Berlin: Suhrkamp, pp. 72–89.
- Pocock, J.G.A. (1975): *The Machiavellian Moment. Florentine Political Thought and the Atlantic Republican Tradition*. Princeton: Princeton UP.
- Rancière, Jacques (1995): *La mésentente. Politique et philosophie*. Paris: Galilée.
- Shakespeare, William (2008): *Macbeth*. Updated edition. Ed. by A.R. Braunmuller. Cambridge: Cambridge UP.





## Ethics and Value Theory



Marcia Baron

# The Distinction between Objective and Subjective Standards in the Criminal Law

**Abstract:** In the criminal law and criminal law theory, claims are often made to the effect that ‘We can’t say *X* without abandoning an objective standard.’ ‘Objective’ is used without much clarity as to what is meant by it. In an effort to shed some light on disputes in philosophy of criminal law and in legal scholarship, and to undermine some common assumptions in criminal law and criminal law theory, I sort through various ways we contrast ‘objective’ to ‘subjective’, with particular attention to ways the contrast is drawn in criminal law. It is common for various senses of ‘objective’ to be conflated, with the result that options are sometimes ruled out, or arguments put forward, based on confusion about what objectivity requires. In the remainder of the paper I focus on the issue of how ‘reasonably believed’ should be understood in self-defense law, where the defendant is required to have *reasonably believed* (or believed on reasonable grounds) that it was necessary to use self-defensive force. I endorse the position taken in *People v Goetz*, rejecting Goetz’s claim that an objective standard of reasonableness would not allow consideration of the “actual circumstances of the particular incident,” and I then try to expand on some matters not addressed by the *Goetz* in its (roughly correct, but somewhat sketchy) account of objective standards.

## 1 Introduction

My focus will be on objectivity and the notion of the reasonable, in connection with law and philosophy of law, and more specifically philosophy of criminal law.<sup>1</sup>

Several years ago I had the good fortune to hold a fellowship<sup>2</sup> that enabled me to step outside the field of philosophy and study law. My focus was criminal

---

1 The notion of the reasonable plays a part in other areas of the law besides criminal law, but it is important not to assume that the same notion of the reasonable is needed or operative (not necessarily the same thing!) in various areas of the law. Whereas the notion of the reasonable needs to be kept distinct from the rational if it is to be helpful for criminal law, the same may well not be true in other areas of law, e.g. tort law. For the same reason, the common use of ‘reasonable’ as roughly equivalent to ‘prudence’ is seriously problematic in the criminal law, but may not be in some other areas of the law. I discuss this in “Reasonableness,” unpublished manuscript.

2 From the College of Arts and Sciences at the University of Illinois at Urbana-Champaign. I’m forever grateful!



law, and it did not take long to see that the concept of the reasonable plays a major role in statutes and rulings, and in the underlying criminal law doctrines.<sup>3</sup> The word ‘reasonable’ often precedes the words ‘belief’ and even more often ‘person’; often it serves to constrain a defense (self-defense and the heat of passion defense, among others). It is also frequently the locus of important – and philosophically intriguing – controversy. Sometimes the pivotal question is whether a belief should have to be reasonable to exculpate; famously, this was the situation in *Morgan*, an English case from 1976.<sup>4</sup> The *Morgan* case concerned rape: the jury instruction had specified that a belief that the complainant was consenting is a complete defense to rape but only if the belief was reasonable; the defendants appealed, challenging the restriction; the House of Lords ruled that the instruction was indeed erroneous.<sup>5</sup> That is, they ruled that a belief that the complainant was consenting is a complete defense whether or not the belief was reasonable.<sup>6</sup>

Sometimes the locus of controversy is not whether there should be a requirement of reasonableness, but how the term ‘reasonable’ should be understood. This is the case in the law of self-defense in the US. To be acquitted on self-defense grounds, the defendant has to have believed *p* (at the time she committed the act), and the belief, moreover, has to have been reasonable. (I use ‘*p*’ rather than spell out the content of the belief, so as not to be distracted just now by the details of what exactly *p* is, and thus to keep the spotlight on the notion of the reasonable.) But just how the term ‘reasonable’ should be understood was, and continues to be, a matter of considerable debate.

The options for understanding ‘reasonable’ in the reasonable belief requirement for self-defense are commonly referred to as the *objective standard* and the *subjective standard*. There is no clear consensus on what these are. Nor is there adequate recognition that the terms are used in conflicting ways. This is probably due in large part to the false impression that it is just obvious how the objective

---

<sup>3</sup> According to George Fletcher, this is not true in Continental law. See Fletcher 1985 and Fletcher 1988, pp. 39–41.

<sup>4</sup> *Director of Public Prosecutions v. Morgan* 2 All ER 347 (1975). For discussion, see, *inter alia*, Duff 1981; Curley 1976; Baron 2001. See too Glanville Williams’ endorsement of the ruling in Williams 1975.

<sup>5</sup> Since then the law has been modified; now belief that the complainant was consenting exculpates only if it was held on reasonable grounds (<http://www.legislation.gov.uk/ukpga/2003/42/contents>, visited on 5 August 2015).

<sup>6</sup> The defendants’ conviction was not affected, however, because it was considered harmless error. It was apparent that the jurors thought the defendants were lying in saying that they believed the victim was consenting.

and the subjective differ. In fact we use the terms in a number of ways, and the different ways we contrast the objective and the subjective are not equivalent.

My first aim in this paper is to sort out various ways we contrast the objective to the subjective, with particular – but not exclusive – attention to ways the contrast is drawn in the criminal law and in criminal law theory.<sup>7</sup> I'll then note some consequences of a failure to keep these distinct (or even to remember that the objective and the subjective are contrasted in all these different ways). The failure gives rise to claims to the effect that we cannot say *X* without giving up on an objective standard, claims that are in turn relied on to argue against proposals to modify or expand a defense. Or, they are relied on to conclude that unless we are willing to ignore the particulars of the case, we will have to adopt a subjective standard. Later in the paper I focus on self-defense and the issue of whether the reasonable belief requirement should be understood as a subjective or as an objective standard, with particular attention to how 'objective standard' should be understood.

## 2 Eight ways the objective and the subjective are contrasted

The first three items on my list form a cluster and serve as a backdrop to (4), (5), and (7), which are of greater importance for my purposes than the first three. I differentiate (1)-(3) not in order to highlight the differences between them but rather to bring out more sharply the differences between, on the one hand, (1) and (2), and on the other, (4), and more generally, to attune our ears to the differences in the way the contrast is drawn.

1. 'Objective' is sometimes used to mean 'how it actually is'. The basic idea in this way of distinguishing the objective from the subjective is that the objective is (to put it as one might in an introductory philosophy text) the way things really are, and the subjective is the way things seem to the perceiver.

2. Only slightly different from (1): 'objective' is used to point to what actually happened, and 'subjective' to point to the subject's experience of the thing in question (and sometimes to the thing in question as experienced by the subject).

3. A somewhat similar contrast – or more accurately, pair of contrasts – is drawn in discussions of criminal attempts. 'Objective' is used in discussions of criminal attempts to highlight either what actually happened, or what the agent in fact was doing. 'Subjective' is used to draw attention to what the

---

<sup>7</sup> Because of my particular interest in criminal law, I don't aim for a complete list of ways we contrast the objective to the subjective.

agent aimed to do, or (and this will sometimes be different) to what she thought she was doing.

The subjective/objective contrast will differ depending on the type of attempt, and how the attempt fails. Consider a pair of examples from Antony Duff: “Subjectively, John shoots Pat (for that is what he intends or expects to do); objectively, his shot misses her. Subjectively, Ms Ryan handled stolen goods (that was what she believed she was doing); objectively, she handled non-stolen goods” (Duff 1996, p. 194).<sup>8</sup> In the first case, the contrast is between what John intended and what happened. In the case of the supposedly stolen goods, the contrast is a bit harder to articulate, but it is not between what she intended and what happened; the gap is not (at least not only) between intention and result, but between *her construal of the action* and *what the action in fact was*. Hence there are two different subjective/objective contrasts here:

3a: ‘Objective’ points to what in fact happened, and ‘subjective’ to what the agent intended;

3b: ‘Objective’ points to what the action in fact was, and ‘subjective’ to how the agent construed it (to what she thought she was doing).

Note that the nature of John’s failed attempt to shoot Pat matters. Had he failed to shoot Pat not because he missed but because what he shot at (and suc-

---

<sup>8</sup> In borrowing Duff’s examples, I am not signing on to what he does with them. He claims that in the law of attempts, there is a “second kind” of objectivity, which concerns what a reasonable person “would believe or expect, as distinct from what this particular agent (perhaps unreasonably) believed or expected. Subjectively, the would-be killer by witchcraft is trying to kill; objectively, he is not, since any ‘reasonable person’ would know that this is not a possible method” (Duff 1996, p. 194). This parallels the contrast I draw below between the sense of ‘objective’ in (4) and that in (1)-(3), although because of the intricacies in attempts – the differences in the various ways one’s attempt, despite one’s earnest efforts, may fail – the objective/subjective contrast is rather different here, and the contrast Duff draws is less sharp than the contrast I draw between ‘objective’ in (4) and ‘objective’ in (1)-(3). Compare the objective/subjective distinction with respect to (a) S’s attempt to kill using witchcraft and (b) S’s failed attempt to fatally shoot someone (failed because her shot misses him). In each case the contrast is between what S aims to do and what in fact happens. In each case S tried to kill someone and failed. It is not clear that the fact that in the first instance S chose a method that could not possibly work (or rather, could not work unless her actions so frightened the person that it was the catalyst to a heart attack) calls for a different way of contrasting the subjective and the objective. In other words, it is not clear that it involves, as Duff says it does, a second kind of objectivity. Nor is it clear that there is any need to make reference to what a reasonable person would have believed. Not, that is, if we are only trying to capture the differences between the two failed attempts. Duff makes reference to it in the service of explaining competing views concerning criminal liability for attempts. In this context, the role of the reasonable person is quite different from the role played in self-defense, provocation, and in definitions of recklessness and negligence.

ceeded in hitting) was a dummy she left in her bed to trick him, we might say (or at least a legal theorist might say) that subjectively he shot Pat but objectively he shot a dummy. If we said this, we would be drawing a rather different subjective/objective contrast than in the case where he aimed at Pat but missed. It would be a case of 3b. (It could count as a case of 3a as well, but a special kind of case; 3b is critical to understanding it.) The contrast in the revised case, where he shot a dummy, but thought that he shot Pat, is less like the contrast where he aimed but missed, and more like that in the case of the supposedly stolen goods.

I won't be discussing attempts, and bring them up mainly to differentiate the uses of 'objective' and 'subjective' (and 'objectively' and 'subjectively') in the context of attempts from the somewhat similar uses of the terms in other criminal law contexts, and thus to forestall confusion.

4. Another way of contrasting the subjective and the objective understands 'subjective' roughly as in (1) and (2), but means by 'objective' something importantly different. To tease this out, consider a common division of the elements of a defense into the subjective and the objective components. The traditional heat of passion defense, for example, is typically explained as having both a subjective and an objective element.

The defense requires that "provocation by the victim caused the accused temporarily to lose self-control". That is the subjective element. The defense also requires that "a reasonable person [...] similarly situated may also have been caused temporarily to lose self-control", and this is routinely referred to as the "objective element" or "objective test" (Klimchuk 1994, p. 441).<sup>9</sup>

The contrast here is between on the one hand, *something that happened to the defendant or something the defendant experienced* (and it could, in a different context, be something the defendant believed or thought or feared), and on the other, *what a reasonable person would, or might,*<sup>10</sup> *have experienced* (or believed or thought or feared). The contrast is not between what the defendant thought (or felt or experienced or saw) and what *actually was the case* (or actually did happen). Rather, it is between the former, and what a reasonable person in the same situation would or might have experienced, thought, feared, etc. The objective component in (4) is, to be sure, a bit mysterious; I can see why some might have qualms about it. Yet it is, I believe, vital to the law.

---

<sup>9</sup> I've taken this from Klimchuk 1994, but it is quite standard and can be found in many sources. Sometimes it is formulated in terms of the ordinary person rather than the reasonable person; Klimchuk presents it both ways.

<sup>10</sup> It is more common to say simply *would* and to say *what the reasonable person would have experienced* (etc). Klimchuk does not, and I think it important not to do so. For more on this, see Baron 2012.

Before proceeding to the fifth distinction, I want to underscore the importance of not conflating the notion of the ‘objective’ in (4) with the notion of the ‘objective’ in (1), (2) and (3). Although there are differences in the notion in (1), (2), and (3), the basic idea is the same there, and sharply different from that in (4). I emphasize this because sometimes the standard of the reasonable person is spoken of as if it involved absence of error, as if the standard were that of perfection (see Byrd 2005). Clearly, this would render the standard unhelpful for purposes of the criminal law.

5. A different way of distinguishing between ‘objective’ and ‘subjective’ is evident in the applications of these terms, in criminal law and legal scholarship, to standards or principles, or to reasonableness as determined by a standard or principle. Note that whereas in (4), the terms ‘objective’ and ‘subjective’ refer to the content of requirements for a particular defense (and one could use them to refer likewise to the content of principles), the idea here is different: the standard itself, not what it is about, is said to be either objective or subjective. I’ll be exploring in Sect. 4 how best to make sense of this distinction, applied to standards of reasonableness.

6. ‘Objective’ and ‘subjective’ are also used in a way that is pointedly evaluative. ‘Objective’ and ‘objectively’ can signify ‘without bias’ or ‘without regard to one’s own interests’. ‘Subjective’ and ‘subjectively’ signify ‘biased’ and operate as terms of negative appraisal. The bias can be in favor of one’s own interests (or those of one’s company or another group of which one is a member), or the interests of particular others. We may also say of someone that she is biased when we think that her likes or dislikes, or her political or religious or other views, inappropriately shape her research, or assignments of grades to her students, or (if she is a judge) her verdicts. (And we might also say of her research or evaluations of her students or judicial decisions that it or they are biased.)

7. Different from (1)-(6) is the following sense of ‘objective’: that a standard is objective is held to preclude taking into account certain particulars of the situation and especially of the defendant, and sometimes is assumed to preclude taking any of the defendant’s personal characteristics into account at all.

That it precludes taking any of the defendant’s personal characteristics into account is implausible, but by no means an unusual assumption. The assumption is in evidence in the following statement, part of an explanation of the heat of passion defense: “As in other areas of the criminal law [...] there is a movement to subjectivize the standard, i.e., to include at least some of the defendant’s personal characteristics in the ‘ordinary/ reasonable person’ standard.” (Dressler 2009, p. 539) The claim that this amounts to “subjectivizing” presup-

poses that if any of the defendant's personal characteristics are included, the standard is not objective, at least not fully objective.<sup>11</sup>

Similarly, an objective standard of reasonableness is explained as follows in *State v. Leidholm*:

An objective standard of reasonableness requires the fact-finder to view the circumstances surrounding the accused at the time he used force from the standpoint of a hypothetical reasonable and prudent person. [...] Ordinarily, under such a view, the unique physical and psychological characteristics of the accused are not taken into consideration in judging the reasonableness of the accused's belief.<sup>12</sup> (*State v. Leidholm* 334 N.W.2d (N.D. 1983), p. 817)

It is not unusual for such statements as the one from *Leidholm* to include 'fully' or 'thoroughly' (or a similar qualifier) in front of 'objective', thus treating objectivity and subjectivity as scalar notions. There is merit in viewing them as not forming a clean dichotomy, but it is a mistake to suppose that the more personal characteristics we allow as relevant – the more that can permissibly be taken into account in applying the standard – the less objective the standard. Nor is it the case that a standard is automatically less than fully objective if any particulars are allowed to factor in.

A more moderate position is much more plausible, namely, that *some* particulars of the defendant (or the situation)<sup>13</sup> are such that taking them into account is incompatible with a standard being objective. The interesting question then is which particulars fit this description. I'll explore this shortly.

8. Worth bearing in mind is a rather different contrast from those drawn above. The idea is that if something is objective—as when people speak of “objective facts”—it is a matter on which disagreement, when everyone looks or reflects with care, should not persist. When disagreement does persist, and we think it is for reasons other than obtuseness, we speak of there being an element

---

**11** Although I have followed Dressler and others in using the word 'personal', I am not sure what work that word is doing, and so do not know whether, on Dressler's conception of subjectivizing, there are characteristics of the defendant that can be included consistent with the standard being objective. My guess is that 'personal' is meant to exclude generic qualities that persons in general have.

**12** As with 'personal' in the quote from Dressler, I am not sure what 'unique' is supposed to add, but here too the idea is probably to differentiate generic human characteristics from those specific to that individual. I don't think the idea is that they have to be unique.

**13** I include 'or the situation' because although the two quotations, from Dressler and from *Leidholm*, concern only particulars about the person, it is not unusual to think that the particulars of the situation are not to be taken into account.

of subjectivity.<sup>14</sup> The dimensions of an object would be classified under ‘objective’; its attractiveness, under ‘subjective’. It is probably this sense of ‘objective’—that the matter is something on which disagreement, when everyone looks or reflects with care, should not persist—that underlies a distinction sometimes drawn between “objective” evidence and “subjective” evidence (an example of the former being lab results, and of the latter, eyewitness reports).

### 3 Connections and conflations

First, a comment on (6). Clearly there is a strong connection between impartiality and this sense of objectivity. When we criticize something for being subjective or for having been done subjectively or if we emphasize the need to be objective, usually the idea is that impartiality is called for. But not always. If someone asks, regarding a third party’s judgment that  $p$ , “How objective is that?”, their worry might not be exactly that the person was failing to be impartial, but rather that he employed, in arriving at  $p$ , a rather relaxed standard of evidence in order to facilitate concluding what he was hoping to conclude. Or (not quite the same, but similar), we may worry that he discounted or dismissed some evidence in order to avoid reaching the conclusion to which the evidence pointed.

More important for my purposes is a connection between (4) and (5). Wariness about objectivity of the sort we see in (4) provides momentum for a subjective standard of reasonableness. Some have grave doubts about holding people to a reasonable person standard and for that reason, if there is a requirement of reasonableness, they champion what they speak of as a subjective standard of reasonableness. In that way, they can dilute the force of the requirement considerably, perhaps in effect rendering it toothless. What a subjective standard amounts to will become a little clearer in the next section, through an examination of the Goetz case.

Before we look at that case, a general observation is in order. Each of (1)-(8) captures something about objectivity. That does not entail—nor is it the case—that for something to be objective, it has to be objective in *all* of these senses. What objectivity requires will vary from one context to another, depending on what sense of ‘objective’ is apt for the particular purposes at hand.

---

<sup>14</sup> In a review of Greenawalt 1992, Neil MacCormick writes that Greenawalt concludes that “on any fine point of balancing, reasonable people can differ. These differences are not objectively corrigible. To that extent, there remains an element of apparently irreducible subjectivity in the inevitable leeways of legal judgement” (MacCormick 1999, pp.1575 – 1604).

I emphasize this because often in discussions in criminal law, it is claimed that to allow *A* would be to give up on an objective standard. Or it is claimed that to adhere to an objective standard precludes taking *B* into account. The background assumption seems to be that because in some contexts objectivity requires *C*, objectivity always requires *C*. Not that if asked, very many people would be likely to affirm this. What may happen, though, is that people latch onto a particular sense of the word ‘objective’ or ‘objectively’ (or an amorphous blend of various senses) and without thinking about it, suppose that objectivity invariably is like that. They then implicitly treat as a test for objectivity something that is in fact not suitable for that context. A related phenomenon is that other senses of ‘objective’ are simply forgotten. Authors claim objectivity to be on their side when it is only in one sense of the term that it is on their side, and in an equally important (but overlooked) sense of the term, objectivity is on the opposing side (or an opposing side).

Consider the following example of this last phenomenon. It is fairly common to hold that mistaken self-defense can at best be excused, never justified, even when the mistake is a reasonable one. This reflects an underlying view that justification requires truth. I don’t want to claim that this position is just wrong or confused, but I think it better to understand justification as requiring reasonable belief rather than truth. The view that it requires truth is fed by the conjunction of the following: the common view that justification is objective, and that (as (8) above suggests), ‘objective’ entails that there is no room for disagreement.

Here is what I have in mind. It is often claimed that it cannot be the case that *A* is justified in attacking *B* in self-defense and that *B* is justified in using force to thwart *A*’s attack; they cannot both be justified. They can both be excused, or one can be justified and the other excused, but (it is said) they cannot both be justified.<sup>15</sup> Why not? Because justification is objective, and ‘objective’ entails that there is no room for disagreement. This is deployed as an argument for the position that mistaken self-defense can at best be excused, and more generally for the position that justification requires truth. This is one of many instances where one sense of ‘objective’ is appealed to as if it provided a necessary condition for something to qualify as objective.

What I described concerning the dispute about whether justification requires truth reflects a broader problem: accepting (8) as a requirement for objectivity in general leads people to neglect, misread, or under-appreciate the importance of the notion of objectivity in (4) and sometimes to either dismiss it or read into the notion in both (4) and (5) the sense of objectivity that we see in (1). That is, there

---

<sup>15</sup> I discuss the claim at greater length in Baron 2005, esp. sect. VIII.



is a tendency to suppose that ‘objective’ in the sense of ‘what a reasonable person might judge/feel’ or ‘how a reasonable person might react’ entails ‘free of error’. The contrast (in 1) between how things actually are and how they might seem to the perceiver is read into (4), with the result that one may forget that a reasonable person is sometimes in error. Likewise with respect to (5): it is all too often forgotten that meeting an objective standard does not require that one not be in error.

Let me pause to take stock. I began by saying that I wanted to look at various ways in which we contrast the objective to the subjective, and I had in mind both how this is done in the criminal law and criminal law scholarship, and how it is done outside of criminal law, in ordinary discourse. My interest in this stems from my puzzlement over some common claims, on which a fair amount rests, concerning objective and subjective standards. I suspect that various senses of ‘objective’ are conflated, with the result that something is thought to be at odds with objectivity or an objective standard, when in fact it is not. What I’ve been talking about is one illustration of that; another will emerge in the next section.

I turn now to the Goetz case. Intriguing for its discussion of the difference between objective and subjective standards, the Goetz case will be helpful for reflecting on questions raised above under (7). My position is that it is a mistake to think that a standard is more objective the fewer features of the situation, or the people involved, it treats as relevant. Nonetheless, it is true of some features (though far fewer than is often thought) that treating them as relevant disqualifies the standard from being objective. The challenge is to work out which features are such that taking them into account is incompatible with an objective standard.

## 4 Goetz and the word ‘reasonably’

In 1985 Bernhard Goetz, a white man, shot four young African-American men when they asked him, in a subway car in New York City, for \$5. Goetz was indicted for attempted murder (among other things). He challenged the indictment, claiming that the instruction given to the (second) grand jury concerning what is meant by ‘reasonably’ was erroneous. Let’s look into this claim.

The NY statute on self-defense states that “a person may [...] use physical force upon another person when and to the extent he reasonably believes such to be necessary to defend himself or a third person from what he reasonably believes to be the use or imminent use of unlawful physical force by such other person” (NY Penal Law 35.15(1). Quoted in *People v. Goetz*, 497 N.E.2d (N.Y. 1986)). The jury instruction to which Goetz objected was provided in response to a juror’s request for clarification of the term ‘reasonably believes.’

The instruction was “to consider the circumstances of the incident and determine ‘whether the defendant’s conduct was that of a reasonable man in the defendant’s situation’”. Goetz claimed that the response provided was erroneous, and that “the introduction of an objective element will preclude a jury from considering factors such as the prior experiences of a given actor and thus, require it to make a determination of ‘reasonableness’ without regard to the actual circumstances of a particular incident.” According to Goetz, a proper instruction would be to consider *not* “whether the defendant’s conduct was that of a reasonable man in the defendant’s situation” but “whether a defendant’s beliefs and reactions were ‘reasonable to him’” (*People v. Goetz*, p. 48). The lower courts sided with Goetz. The highest court of the state rejected his argument and reinstated the indictment. I see little to say in favor of the standard Goetz proposed; clearly ‘reasonable’ is doing virtually no work at all if we take the standard to be asking whether the defendant’s beliefs and reactions were “reasonable to him.” Although it is worth considering whether there might be a more plausible subjective standard,<sup>16</sup> here I focus on Goetz’s argument against the jury instruction.

We can see in his argument the notion of ‘objective’ in (7): objectivity is thought to be at odds with taking into account various particulars, in this instance, the prior experiences of the defendant and the actual circumstances of the incident. Why might one think these are at odds? There appears to be a background assumption that taking an objective view means taking a highly abstract view. That is not a pure fabrication; there may be such a notion of objectivity. Thomas Nagel wrote in *The View from Nowhere*:

[T]he distinction between more subjective and more objective views is really a matter of degree, and it covers a wide spectrum. A view or form of thought is more objective than another if it relies less on the specifics of the individual’s makeup and position in the world, or on the character of the particular type of creature he is. The wider the range of subjective types to which a form of understanding is accessible—the less it depends on specific subjective capacities—the more objective it is. A standpoint that is objective by comparison with the personal view of one individual may be subjective by comparison with a theoretical standpoint still farther out. The standpoint of morality is more objective than that of private life, but less objective than the standpoint of physics. (Nagel 1986, p. 5)

This way of thinking about objectivity may have its place. But clearly, the sense of ‘objective’ in which the jury instruction is putting forward an objective standard is that articulated in (4) above, not the one articulated by Nagel. It seems likely that a failure to differentiate various senses in which we call something ‘objective’ underlies an assumption that I noted above in discussing (7), namely that a standard is

---

<sup>16</sup> I do so in a work in progress, *Self-Defense, Reason, and the Law*.

objective to the extent that it abstracts from the particulars. The contrast between the subjective and the objective that Nagel articulates makes some sense as applied to *standpoints*, but not, as far as I can see, to *standards*.

That said, (7) is not irrelevant to (4), and the moderate position I mentioned in explaining (7) – that some particulars of the defendant or the situation are such that taking them into account is incompatible with a standard being objective – is needed to make sense of the idea of a reasonable person standard. A reasonable person standard asks us to consider whether a reasonable person in the defendant's situation might have acted as the defendant did; and the question then arises, what counts as the defendant's situation? If we include as part of the situation everything that is true of the defendant, then the standard does little more than ratify whatever the defendant did. (Might a reasonable person who was exactly like the defendant have acted as the defendant did? Of course!) The reasonable person standard has to treat as *not* part of the situation *some* of the defendant's qualities. But which ones?

The New York Court of Appeals (the highest court in the state of New York) spoke to this issue. Rejecting Goetz's argument and reinstating all counts of the indictment, the court wrote that his argument

falsely presupposes that an objective standard means that the background and other relevant characteristics of a particular actor must be ignored. To the contrary, we have frequently noted that a determination of reasonableness must be based on the 'circumstances' facing a defendant or his 'situation.' Such terms encompass more than the physical movements of the potential assailant. [T]hese terms include any relevant knowledge the defendant had about that person—including, for example, any information the defendant had concerning the assailant's prior acts of violence or reputation for violence. They also necessarily bring in the physical attributes of all persons involved, including the defendant. Furthermore, the defendant's circumstances encompass any prior experiences he had which could provide a reasonable basis for a belief that another person's intentions were to injure or rob him or that the use of deadly force was necessary under the circumstances. (*People v. Goetz*, p. 52)

I think the NY Court of Appeals had the right idea. The objectivity of a standard is not impugned if we take into account prior experiences that gave the defendant reason to believe that the person intended to injure or rob him or that retreat would not be safe and that it would be necessary to use force to thwart the attack. Likewise, there is no reason to think that an objective standard, in virtue of being an objective standard, rules out treating as part of the "situation" or "circumstances" the size and weight of the apparent assailant (hereafter *A*) relative to those of the defendant (hereafter *D*), and information that *D* had about *A*'s prior acts of violence. This is a very important corrective. There is no reason to think that a standard is less objective if it allows us to take these features into account.

But other features *are* ruled out, and the NY Court of Appeals hints by omission at what these are: *D*'s beliefs<sup>17</sup> and attitudes, and prior experiences of *D*'s that do not provide a reason to believe that the person intends to injure her and that it is necessary to use force to thwart the attack. Although the Court does not explicitly state that a standard cannot count as objective if it treats these as part of *D*'s situation, this clearly is the suggestion. And the suggestion is certainly plausible, with one minor qualification: *D*'s beliefs and attitudes could be relevant insofar as *D* has reason to believe that *A* hates people who have (say) certain political beliefs. That *D* has those belief(s) would then be relevant to her assessment that *A*'s threat against her is to be taken very seriously (though really what matters is not so much that she has these political beliefs as that *A* believes she does). That is the only way in which the defendant's beliefs and attitudes can be relevant, on an objective standard.

To delve further into which features of the defendant or the situation are, and which are not, such that an objective standard can treat them as relevant, we need to focus our attention on the Court's suggestion that prior experiences can be relevant insofar as they "provide a reasonable basis for a belief that another person's intentions were to injure or rob him or that the use of deadly force was necessary under the circumstances". Clearly there is room for disagreement concerning what sorts of experiences do provide that basis.

Two considerations should guide us here. First, an experience need not, all by itself, constitute adequate reason for believing *p* in order to count as providing a reasonable basis for a belief that *p*.<sup>18</sup> (I use '*p*', as I did earlier, as shorthand for the content of what one needs to believe to have a self-defense claim.) The bar is set too high if we require that it constitute an adequate reason. Second, the mere fact that an experience (more precisely, that the fact that the person had that experience) renders more understandable his believing *p* does not suffice to establish that the experience provides a reasonable basis for his belief that *p*.

My view – perhaps an unusual one – is that experiences of having been the victim of a crime on a previous occasion (even a victim of a crime in similar circumstances, e. g., while in a subway train) would *not* generally provide a reasonable basis for the belief that this person was about to attack one. Sometimes they would. They obviously would, first of all, if the person who attacked one before is (or is plausibly believed by the agent to be) the same person who is acting suspiciously towards one now, or is affiliated (or plausibly believed to be affiliated)

---

<sup>17</sup> Not to be confused with information the defendant has that provides a basis for her belief.

<sup>18</sup> I thank the students in Gabe Mendlow's seminar at the University of Michigan Law School, especially Mohammad Alharoun, for leading me to see this.

in some significant way with that person (e.g. is hired by the other person to carry out violent actions on his behalf). In addition, they might do so – providing less of a basis, but nonetheless some – if the prior experiences taught one something about how these attacks are often carried out. But if we bracket cases where the same person who subjected one to violence before is now (at least seemingly) attacking one or threatening to attack, prior victimization very often does not provide a reasonable basis for a belief that *p*. The relevance of prior victimization to the reasonableness of one's belief – on an objective standard of reasonableness – tends to be exaggerated.

The first point requires little explanation: if *A* has struck *B* in the past, often very soon after accusing her of smiling flirtatiously at another man, if *A* tends to strike *B* especially violently when drunk, sometimes then hiding *B*'s insulin, once thereby causing *B* to go into a diabetic coma,<sup>19</sup> and if *A* now seems angrier than in the past and is threatening to kill *B*, undeniably the prior experiences *B* has had with *A* give her added reason to think that her life is in danger if he is now drunk and accusing her of smiling flirtatiously at another man. These experiences (both of having been the victim of his violence before, and specifically having been repeatedly victimized by him in circumstances relevantly similar to those she is in now) provide a reasonable basis for a belief that she needs to use force to save her life. So do other experiences with him that equip her to judge how seriously angry he is with her. (How strong a basis they provide depends on other factors, among them, her options of escape; if they are getting a divorce, and not against *A*'s wishes, and *A* – very atypically for abusive men – just wants to wash his hands of *B* and will gladly let her disappear from his life, perhaps she can safely flee. But typically, flight from a violent, abusive male partner only intensifies his anger at her.<sup>20</sup> Also relevant is their relative size and fighting abilities.)

The second is more complicated. Suppose the very same *B* succeeds in leaving *A*. Two years later, in a new relationship, she is accused by *C*, her new partner, of flirting with another man. *C* is angry. He glowers at her, and she smells alcohol on his breath. He has never struck her or been violent towards her in any other way; he does not threaten her now; she has never heard that he has a tendency to be violent. She is in the kitchen when he accuses her; he is larger than she is; he is between her and the exit from the kitchen; she feels trapped. She grabs a large knife that is at hand and stabs him. Does the experience of being abused by *A* provide a reasonable

---

<sup>19</sup> For a case where the abuse included hiding insulin, see *State v Hundley*, 693 P.2d 475 (Kan. 1985). I discuss the case in Baron 2011.

<sup>20</sup> See, *inter alia*, Mahoney 1991.

basis for her belief (or perhaps a more apt word is ‘feeling’ or ‘sense’) that she is in danger now and needs to use lethal force to protect herself? No. It renders it far more understandable than it would otherwise be, and has some exculpatory significance. It would certainly be relevant to bring up at sentencing, and ideally there would be an excuse defense to which it would be relevant (a partial excuse, rather than one that would result in acquittal).<sup>21</sup> But it does not provide a reasonable basis for her belief.

Could any experience of prior victimization provide a reasonable basis for a belief that *p*, other than when the current (apparent) attacker is the same person as the previous attacker, or a close associate of that person? Not often. But it might, at least if the approach (gestures, words used, positioning of the persons approaching *D*) is so similar to that of the aggressor(s) who attacked *D* before, and so out of the ordinary, that it seems clear that they are employing the same methods.

Consider the following example (put to me as evidence that prior victimization can indeed provide a rational basis for a belief that *p*): on a previous occasion, *D* was asked in a relatively empty subway car for money by a young man *D* did not know, accompanied by three others; *D* tried to appease them by handing them a few dollars, but to no avail; they attacked *D*, causing moderate though not grave harm. Thereafter *D* carried a loaded gun when he rode the subway. Two years later he is again approached by young men who ask him for money, in a manner that resembles that of the (previous) attackers. The question is: does the prior experience provide a reasonable basis for his belief that he needs to use force to protect himself? Maybe. For although this case is not entirely unlike that of the battered woman who stabs her new partner in that in both cases, the action taken might well not be followed by violence, it differs in that we think the subway request (given the positioning of the four youths) more likely to be followed by violence.

The reason I say it is easy to exaggerate the relevance of prior experience to the reasonableness of the belief is this: We all know that a group of young males asking someone in a subway for money might be planning to attack him or her; we do not need to have experienced an attack to know that. We also all know (or should!) that they might be happy just to receive a very small amount of money, and might also very well leave one unharmed even if one gives them nothing. What does the experience of victimization add to one’s knowledge? Does one really have more by way of a reasonable basis for a belief in the need to use

---

<sup>21</sup> In England the heat of passion defense has been replaced with a defense that might arguably apply to this sort of case (see <http://www.legislation.gov.uk/ukpga/2009/25/sections/54-56>, in particular 55 (3), visited on 5 August 2015).

self-defensive force if one has been victimized in the past than if one hasn't?<sup>22</sup> It is not clear to me that one does. The experience explains why the danger was so very salient to *D*, and helps to explain why *D* shot the four youths, but it is not evident that it provides more by way of a reasonable basis for one's belief that *p* than he would have had in the absence of prior victimization. We might say, too, that prior victimization gives one a more vivid knowledge. But rendering something more salient, or yielding more vivid knowledge, is not the same as providing a reasonable basis for a belief.<sup>23</sup>

I have dwelled on this at some length because it is so common to treat prior victimization as very frequently, perhaps typically, providing a reasonable basis for a belief that one is now about to be attacked. In a case such as the one I described – where the situation in the prior victimization was very similar to the present one – the matter should be put to jurors, so that they can assess whether the defendant had more of a basis for a belief that *p* than she or he would have had in the absence of prior victimization. But at the same time, jurors should be instructed that the fact that the prior victimization would make the use of (lethal) force more understandable does not support the claim that the belief was reasonable.

## 5 Concluding remarks

I have sought in this paper to distinguish various ways we differentiate the objective from the subjective, in order to draw attention to ways in which conflating them has contributed to confusion and error in criminal law. I hope that my list, by no means purporting to be complete, will both encourage more care in thinking about the distinction between the objective and the subjective and especially between objective and subjective standards, and prompt some others to add to my list other examples of conflation in criminal law discussions of different senses of 'objective' and of 'subjective'.

In addition, I have defended the claim that objective standards are compatible with taking into account the relevant particulars that need to be taken into account in self-defense. I have endorsed (with a minor qualification) the *Goetz*

---

<sup>22</sup> I am assuming throughout this paragraph and the next that the attacker is not the same person (or close associate) of the previous attacker, or believed to be by *D*.

<sup>23</sup> I am very grateful to Gabe Mendlow and his seminar students at the University of Michigan Law School for stimulating discussion on the points in the preceding two paragraphs. The example is a modified version of an example put to me by Mohammed Alharoun, in comments on the draft of my paper that we discussed in Prof. Mendlow's class. Special thanks to Alharoun, but thanks to all the law students who provided me with comments.

Court's explanation of an objective standard, and I have tried to expand, where the Court did not, on what sorts of experiences should count as providing a reasonable basis for a belief that  $p$ . One might respond, though, that if the bar for counting as a reasonable basis for a belief that  $p$  is going to be set as high as I have set it, it would be preferable either to reconceive what counts as an objective standard or favor a subjective standard. It might seem unfair to the defendant to use an objective standard of reasonableness for self-defense, yet at the same time to understand 'objective standard' in the way I am proposing.

In brief reply: it should not seem unfair as long as we bear in mind that the defendant is not judged solely by reference to an objective standard of reasonableness, though she is so judged for purposes of self-defense. Factors that cannot enter in for those purposes are not ignored; they can be taken into account both for purposes of any excuse defense that may be available, and as extenuating factors to consider at sentencing. That said, one might wish to develop a more plausible subjective standard than the one Goetz put forward, and might do so by adding (judiciously, and sparingly) to what can count as part of the defendant's situation more than the *Goetz* court allowed. But as long as self-defense remains (as I hope it will) a justification, rather than an excuse, an objective standard of reasonableness seems to be what is called for. And because the reasonable and the understandable are not identical, it is important not to incorporate into the defendant's situation factors that merely make her conduct understandable, but do not go towards making it reasonable. Factors that make it understandable should be relevant only to excuse defenses, and to the post-conviction question of what sentence is appropriate.<sup>24</sup>

## Bibliography

Baron, Marcia (2001): "I Thought She Consented". In: *Noûs* 35, pp.1–32.

Baron, Marcia (2005): "Justifications and Excuses". In: *Ohio State Journal of Criminal Law* 2, pp. 387–406.

---

<sup>24</sup> Earlier versions of this paper were presented in 2014 at the 37th International Wittgenstein Symposium, Kirchberg am Wechsel, Austria, and at the University of Michigan Law School. I am grateful to organizers and discussants at the International Wittgenstein Symposium, to Gabe Mendlow for inviting me to share my work with his students at the University of Michigan, and to both him and his students for their written and oral comments. I would also like to thank the National Endowment of the Humanities for fellowship support in 2010, when I began work on the project of which this forms a part, and Frederick F. Schmitt for helpful discussion of the topics in this paper.



- Baron, Marcia (2011): "Self-Defense: The Imminence Requirement". In: Leslie Green/ Brian Leiter (Eds.): *Oxford Studies in Philosophy of Law*, vol. 1. New York: Oxford University Press, pp. 228–266.
- Baron, Marcia (2012): "The Standard of the Reasonable Person in the Criminal Law". In: RA Duff/Lindsay Farmer/SE Marshall/Massimo Renzo/Victor Tadros (Eds.): *The Structures of Criminal Law*. Oxford: Oxford University Press, pp. 11–35.
- Byrd, B. Sharon (2005): "On Getting the Reasonable Person Out of the Courtroom". In: *Ohio State Journal of Criminal Law* 2, pp. 571–577.
- Curley, Edwin (1976): "Excusing Rape". In: *Philosophy and Public Affairs* 5, pp. 325–360.
- Dressler, Joshua (2009): *Understanding Criminal Law*, 5<sup>th</sup> ed. Newark, New Jersey: Lexis Nexis Matthew Bender.
- Duff, RA (1981): "Recklessness and Rape". In: *Liverpool Law Review* 3, pp. 49–64.
- Duff, RA (1996): *Criminal Attempts*. Oxford: Clarendon Press.
- Fletcher, George P. (1985): "The Right and the Reasonable". In: *Harvard Law Review* 98, pp. 949–982.
- Fletcher, George P. (1988): *A Crime of Self-Defense: Bernard Goetz and the Law on Trial*. Chicago: University of Chicago Press.
- Greenawalt, Kent (1992): *Law and Objectivity*. New York: Oxford University Press.
- Klimchuk, Dennis (1994): "Outrage, Self-Control, and Culpability". In: *University of Toronto Law Journal* 44, pp. 441–468.
- MacCormick, Neil (1999): "Reasonableness and Objectivity". In: *Notre Dame Law Review* 74, pp. 1575–1604.
- Mahoney, Martha R. (1991): "Legal Images of Battered Women: Redefining the Issue of Separation". In: *Michigan Law Review* 90, pp. 1–94.
- Nagel, Thomas (1986): *The View from Nowhere*. New York: Oxford University Press.
- Williams, Glanville (1975): "Lords' Decision on the Law of Rape". In: *London Times*, May 8, 1975, at 15.

Inga Römer

## Gibt es einen kantianischen Intuitionismus in der Ethik?

**Abstract:** Is There a Kantian Intuitionism in Ethics? Within the renaissance of ethical intuitionism, there is one line of argument that wants to avoid the problems with classical intuitionism by giving intuitionism a Kantian form. On the other hand, there are interpretations of the Kantian texts, that read Kant himself as an intuitionist. In this double sense, there is currently a “Kantian intuitionism in Ethics” and the first two parts of the paper reconstruct its particular form in each case. In a third part however, the paper argues, that these two options are problematic and that their general goal can be realized better within a Kantian phenomenology of the Ethical, that would be systematically more convincing and historically closer to the argument developed by Kant himself.

Zu Beginn des zwanzigsten Jahrhunderts hatte der ethische Intuitionismus sowohl in der analytischen als auch in der kontinentalen, und dabei insbesondere in der phänomenologischen Philosophie starke Fürsprecher. Auf der einen Seite stehen Henry Sidgwick, George Edward Moore, Harold Arthur Prichard und William David Ross, auf der anderen Max Scheler, Edmund Husserl, Nicolai Hartmann, Dietrich von Hildebrand und Hans Reiner. Es verbindet sie die Auffassung, dass ethische Urteile nicht aus einem obersten Prinzip abgeleitet werden können, sondern in selbstevidenten ethischen Intuitionen, einem Wertgefühl oder einer Wertnehmung fundiert sind, die sich auf eine Form von moralischen Tatsachen richten, was den ethischen Intuitionismus mit einem moralischen Realismus verknüpft. Uneinigkeit besteht jedoch von Anfang an darüber, was jene moralischen Tatsachen sind, so dass der ethische Intuitionismus bei Sidgwick und Moore eine konsequentialistische, bei Prichard und Ross eine deontologische und bei den Phänomenologen eine wertethische Ausgestaltung anzunehmen vermag. Das Unbehagen mit dem ethischen Intuitionismus hat darin seinen Ursprung, dass er die Ethik auf letzte, selbst nicht mehr begründbare Intuitionen und ihnen korrespondierende moralische Tatsachen zurückführt, in Bezug auf die es strittig ist, ob sie Handlungskonsequenzen, Pflichten oder Werte sind, welchen Status sie haben und wie sie inhaltlich zu bestimmen wären. Sowohl die angelsächsische als auch die phänomenologische Tradition hat seit der Kritik vor allem durch Strawson, Mackie und Rawls von den konsequentialistischen und deontologischen Intuitionismen einerseits und seit Heidegger, Sartre und Levinas von den wertethischen Intuitionismen andererseits Abstand genommen.

Seit einigen Jahren jedoch erlebt der ethische Intuitionismus eine gewisse Renaissance. Diese kann vielleicht darauf zurückgeführt werden, dass der Bereich der angewandten Ethik stetig anwächst und die Aufmerksamkeit auf diffizile Einzelfälle lenkt, in denen universale Prinzipien nicht anwendbar zu sein scheinen. Im Rahmen dieser Renaissance des ethischen Intuitionismus findet sich *ein* Entwicklungsstrang, der den Schwierigkeiten des klassischen Intuitionismus dadurch zu entkommen sucht, dass er ihn mit Grundelementen der kantischen Ethik verknüpft. Diese Entwicklung ist nicht zuletzt deshalb interessant, weil es umgekehrt innerhalb der gegenwärtigen Kant-Forschung Interpreten gibt, die Kant als Intuitionisten und moralischen Realisten verstehen und seinem Formalismus ein intuitionistisches und realistisches Fundament zu geben suchen. Von zwei Seiten her scheint sich damit derzeit ein kantianischer Intuitionismus beziehungsweise ein intuitionistischer Kantianismus anzudeuten. Philosophiehistorisch ist unsere Titelfrage daher eindeutig positiv zu beantworten: Es *gibt* einen kantianischen Intuitionismus. Im Folgenden sei jedoch der Frage nachgegangen, ob und inwiefern das, was unter diesem Namen auftritt, tatsächlich ‚kantianisch‘ und ‚intuitionistisch‘ genannt werden kann und ob es systematisch zu überzeugen vermag. Ein erster Abschnitt skizziert die Gestalt des kantianischen Intuitionismus innerhalb der intuitionistischen Tradition. Ein zweiter Abschnitt zeigt, in welcher Weise Kant als Intuitionist und Realist interpretiert wird. Ein dritter Abschnitt sucht aufzuweisen, dass das von den neueren Bestrebungen Gesuchte sowohl systematisch überzeugender als auch philosophiehistorisch angemessener in einer kantianischen Phänomenologie des Ethischen gefunden werden kann, die sich nur noch in einem sehr eingeschränkten Sinne als Intuitionismus bezeichnen ließe.

## 1 Kantianischer Intuitionismus in der intuitionistischen Tradition

Innerhalb des klassischen Intuitionismus haben Prichard und Ross eine deontologische Variante des Intuitionismus entwickelt, wobei Ross die einflussreichere Figur geworden ist. In seinem 1930 erschienenen Buch *The Right and the Good* vertritt Ross die These, dass das Ethische in anhand von konkreten Handlungssituationen intuitiv eingesehenen, selbst-evidenten ‚*prima facie* Pflichten‘ gründet, die Ausdruck der realen moralischen Ordnung des Universums, gar jedes möglichen Universums seien (Ross 2009, p. 29). Ohne Vollständigkeit oder Letztgültigkeit zu beanspruchen, nennt er sieben derartige Pflichten: Treuepflichten, Reparationspflichten, Dankbarkeitspflichten, Gerechtigkeitspflichten, Wohltätigkeitspflichten, Selbstvervollkommnungspflichten und Nichtverlet-

zungspflichten (Ross 2009, p. 21). Konflikte zwischen diesen Pflichten sind Ross zufolge durch eine an die aristotelische *phronesis* angelehnte praktische Weisheit zu entscheiden (Ross 2009, p. 42). Ross' Position hat insofern einen kantianischen, deontologischen Zug, als sie gegen den konsequentialistischen Intuitionismus von Moore vertritt, Handlungen seien aufgrund ihrer Handlungsart selbst richtig und geboten und nicht wegen der guten Konsequenzen, die sie hervorbrächten. Kant allerdings wirft er eine Überakzentuierung des Motivationsaspekts vor, da Handlungen zwar aufgrund ihres Motivs gut seien, ihre von der Güte ablösbare Richtigkeit jedoch dem verdankten, was durch sie getan und in der Welt erreicht werde. Es kommt Ross dabei stets auf den „höchst persönlichen Charakter der Pflicht“ und damit die Vielfalt der intuitiv erfassten Pflichten an, wobei er es offen lässt, ob sich diese Pflichten nicht doch nach gründlicher Reflexion in einem Prinzip vereinheitlicht fassen ließen (Ross 2009, p. 22).<sup>1</sup>

Diese Offenheit der Ross'schen Position sowie das Unbehagen mit der als intuitiv erfassbar behaupteten Pluralität von inhaltlich bestimmten *prima facie* Pflichten hat Robert Audi im Zuge der Renaissance des ethischen Intuitionismus in seinem 2004 veröffentlichten Buch *The Good in the Right. A Theory of Intuition and Intrinsic Value* (Audi 2004) dazu geführt, den bereits bei Ross angelegten kantianischen Zug zu einem „value-based Kantian intuitionism“ (Audi 2004, p. 200) zu vertiefen. Wie Ross geht Audi von einer Pluralität selbst-evidenter Pflichtprinzipien aus, meint jedoch, dass diesen eine *zusätzliche* Rechtfertigung gegeben werden kann, wenn man sie vereinheitlicht. Als Vereinheitlichungsprinzipien der von Ross angeführten Pflichtprinzipien kommen Audi zufolge insbesondere der kategorische Imperativ in seiner Mensch-Zweck-Formel sowie der Wert des Menschen als eines Würde habenden Zweckes in sich selbst in Frage. In Anknüpfung an einen zentralen Begriff von Rawls vertritt Audi die Auffassung, dass die *prima facie* Pflichten, der kategorische Imperativ und der absolute Wert der Person in einem „reflective equilibrium“ (Audi 2004, p. 144) zueinander stünden, innerhalb dessen sie sich wechselseitig zu stützen, aber auch zu korrigieren vermögen. Dass die *prima facie* Pflichten unter den kategorischen Imperativ subsumierbar seien, gebe sowohl ihnen als auch dem kategorischen Imperativ eine zusätzliche Rechtfertigung; sowohl die Pflichten als auch der kategorische Imperativ aber könnten im Zuge dieses Überlegungsgleichgewichts auch korrigiert und reformuliert werden. Ein gewisses Primat schreibt Audi dem intuitiv erkennbaren, primären Wert der Würde der Personen zu, insofern dieser als der „gemeinsame

---

1 „Loyalty to the facts is worth more than a symmetrical architectonic or a hastily reached simplicity. If further reflection discovers a perfect logical basis for this or for a better classification, so much the better.“ (Ross 2009, p. 23)

Grund“ des kategorischen Imperativs und der Ross’schen Pflichtprinzipien gelten könne (Audi 2004, pp. 142, 112, 111). Allerdings sei dieser primäre Wert der Würde der Person eine „open-ended notion“, der stets nur, dabei jedoch niemals vollständig, durch jene *prima facie* Pflichten, deren Erfüllung zur ihrer Wahrung beiträgt, ein Inhalt verliehen werden könne (Audi 2004, pp. 157–158).

In seiner Anfang 2014 erschienenen Habilitationsschrift (Heinrichs 2013) sucht Bert Heinrichs eine „Ethik der Person“ zu entwickeln, die sich ebenfalls als eine kantianische Variante des ethischen Intuitionismus begreift. Im Unterschied zu Audis Modell des Überlegungsgleichgewichts vertritt Heinrichs jedoch einen ontologischen Minimalrealismus, in dem Personen in ihrer Selbstzweckhaftigkeit die einzigen moralischen Tatsachen sind und nur sie als solche epistemisch intuitiv erfasst werden können.<sup>2</sup> Intuitionismus und Realismus seien ausschließlich auf Personen zu beziehen. Konkrete inhaltliche Prinzipien wie die Ross’schen hingegen müssten durch ein ethisches Konstruktionsverfahren, das sich an dem Interpretationsrahmen der kantischen Selbstzweckformel orientiert, sowie in einer immer wieder neu ansetzenden Hermeneutik der Lebensverhältnisse hervorgebracht werden (Heinrichs 2013, pp. 253, 257). Der inhaltliche, konstruktive Teil der Ethik sei damit prinzipiell revidierbar im Rahmen des Minimalrealismus und Intuitionismus der Person als Selbstzweck. Heinrichs vertritt damit eine hybride Theorie, deren Kern in einem ethischen Intuitionismus sowie moralischen Realismus der Person besteht und deren Zusatz ein ethisches Konstruktionsverfahren vorsieht. Der Intuitionismus ist Heinrichs zufolge durchaus die angemessene ethische Position, man dürfe nur den Rechtfertigungsprozess nicht zu früh – etwa bei den Ross’schen Prinzipien – abbrechen, sondern müsse bis zu dem wirklich realen und intuitionistischen Fundament – der Person – vordringen.

Während Audi meint, dass nichts Wichtiges der kantischen Ethik durch seinen kantianischen Intuitionismus falsifiziert sei, ist Heinrichs sogar der Auffassung, dass seine Position als die kantische verstanden werden könne, insofern sich Kants Ethik „auf ontologischer Ebene nicht nur als Form des Realismus deuten, sondern [...] auf epistemologischer Ebene auch als Variante des rationalen Intuitionismus“ (Heinrichs 2013, p. 263) auffassen lasse. Heinrichs knüpft damit direkt und ausdrücklich an diejenigen Kant-Forscher an, die eine realistische Lesart der kantischen Ethik befürworten.

---

<sup>2</sup> Siehe den Vergleich mit Audi in Heinrichs 2013, Kapitel 6.4.

## 2 Die Interpretation Kants als Intuitionist und moralischer Realist

In kritischer Reaktion auf John Rawls' Auslegung, der von „Kant's Moral Constructivism“ (Rawls 2000, p. 237) spricht, hat sich in der Kant-Forschung eine realistische und intuitionistische Lesart der kantischen Ethik entwickelt. Vertreter dieser realistischen Lesart sind etwa Allen Wood, Paul Guyer, Dieter Schönecker und Alison Hills. Es sind insbesondere zwei Momente in Kants Ethik, auf die sich diese Interpretationslinie stützen kann: ein vermeintlicher Realismus und Intuitionismus des Menschen als eines Zweckes an sich selbst sowie ein vermeintlicher Realismus und Intuitionismus des moralischen Gesetzes.

In ihrem gemeinsam verfassten Kommentar zur *Grundlegung* formulieren Schönecker und Wood: „Bis zum heutigen Tage wird Kants Ethik als deontologische Ethik charakterisiert. Wenn es wesentlich für eine solche Ethik ist, daß in ihr substantielle (an sich existierende) Werte oder Zwecke keine oder bestenfalls nur eine untergeordnete Rolle spielen, dann ist Kants Ethik aber nicht nur nicht deontologisch, sondern dezidiert anti-deontologisch. Denn sowohl die inhaltliche Bestimmung moralischer Pflichten wie auch die Begründung ihrer Gültigkeit hält Kant ohne einen substantiellen Wertbegriff für unmöglich. Vernünftige Wesen als zwecksetzende und autonomiebegabte Wesen haben einen absoluten Wert (Würde); *das* und nicht der Gedanke einer formalen Maximenuniversalisierung ist die zentrale These in Kants Ethik. Daher sind letztlich auch nicht deontische Begriffe für Kant entscheidend (also Begriffe wie ‚verboten‘, ‚geboten‘ und ‚erlaubt‘), sondern Wertbegriffe.“ (Schönecker/Wood 2007, p. 142) Der substantielle, absolute Wert des Menschen als eines Zweckes an sich selbst sei das Fundament der kantischen Ethik. Die beiden Autoren bedauern es, dass Kant dieser seiner Grundthese, der Mensch als vernünftiges Wesen sei Zweck an sich und absolut wertvoll, keine Erklärung hinzugefügt habe, „was es heißt, daß etwas wertvoll ist und wie Werte erkannt werden“ (Schönecker/Wood 2007, p. 207). Hier könnte der Sache nach einer Form von Intuition eine einschlägige Funktion zukommen, wengleich die beiden Autoren Kant in dieser Hinsicht nicht eindeutig einen ethischen Intuitionismus zuschreiben.

In Bezug auf das moralische Gesetz wird Kant ebenfalls ein Realismus und eindeutiger auch ein Intuitionismus zugeschrieben. Allerdings gibt es hier eine Zweideutigkeit in den Auslegungen, die auf die Unterscheidung zwischen Inhalt und Geltung des Gesetzes zurückzuführen ist. Robert Stern spricht von einem Realismus des Gesetzes bei Kant, insofern Kant in mehreren Ableitungen des Gesetzesinhalts zeige, dass das Gesetz nur ein einziges sein und nicht beliebig ausfallen könne. Die Verbindlichkeit des Gesetzes hingegen sei bei Kant kon-

struktivistisch zu verstehen, weil sie durch die Bestimmung der Willkür durch reine praktische Vernunft allererst hervorgebracht werde (Stern 2012, 1. Teil). Er setzt den Konstruktivismus damit früher an als Heinrichs, wenn er die Gesetzesgeltung selbst und nicht erst die konkreten inhaltlichen Prinzipien als konstruiert ansieht. Dieter Schönecker hingegen spricht von Kants ethischem Intuitionismus, insofern die Geltung des einzig möglichen Gesetzes nur intuitiv, und Schönecker meint damit ‚durch die Achtung für das Gesetz‘, erfasst werden könne (Schönecker 2013). Schönecker zufolge vertritt Kant im Rahmen der Lehre vom Faktum der Vernunft einen ethischen Intuitionismus, wobei sich die Intuition der Achtung gleichsam auf die ‚Realität‘ eines wirklich geltenden Gesetzes richtet, während Stern bei Kant zwar einen Realismus des Gesetzesinhalts erblickt, die Gesetzesgeltung jedoch konstruktivistisch auslegt.

Nach dieser skizzenhaften Benennung der Grundgedanken jener realistischen und intuitionistischen Lesart Kants, der, wie oben gezeigt, eine bedeutsame Rolle in der gegenwärtigen Renaissance des ethischen Intuitionismus zukommt, sei nun vor dem Hintergrund von Kants Ethik einerseits und der Phänomenologie andererseits der Ansatz einer kantianischen Phänomenologie des Ethischen umrissen, der sowohl systematisch als auch philosophiehistorisch eine überzeugendere Perspektive anzeigen könnte als die angeführten intuitionistischen und realistischen Konzeptionen.

### 3 Kantianische Phänomenologie des Ethischen

Insbesondere in der Gestalt, die sie bei Heinrichs annimmt, ist die kantianische Wende des Intuitionismus ein deutlicher Fortschritt gegenüber älteren Intuitionismen, die, wie noch Ross, eine Pluralität inhaltlich bestimmter Pflichten zu letzten moralischen Tatsachen erklären. Dem Hang zum Konservativen und zum Ideologischen, der diesen Ansätzen zwangsläufig innewohnte, kann mit der These, allein die Person als Zweck an sich sei eine letzte moralische Tatsache, entgegengewirkt werden. Es fragt sich allerdings, ob nicht auch in diesem Minimalrealismus noch ein *Minimaldogmatismus* steckt. Der kantianische Intuitionismus von Heinrichs nimmt an, dass die Person ontologisch eine Entität ist, die die „*Eigenschaft*“ hat, Subjekt des moralischen Gesetzes und damit Teil einer höheren Ordnung zu sein, an der sie „partizipiere“; diese moralische Tatsache werde von dem entsprechend entwickelten und gebildeten Menschen dann intuitiv erfasst (Heinrichs 2013, pp. 232, 229). Die Person wird hier als eine spezifische ontologische Entität mit spezifischen Eigenschaften aufgefasst, die als moralische Tatsache innerhalb einer nicht-naturalistischen Welt vorkommt, welche von uns unter bestimmten Bedingungen intuitiv erfasst werden kann. Der Minimaldogmatismus dieser Position liegt in der platonisierenden

Annahme einer real existierenden Verstandeswelt: Es gibt real diese höhere Welt der Personen, und unter günstigen Bedingungen erfassen wir sie auch, und zwar intuitiv. Diese zu einer „intuitiven“ intellektuellen Anschauung einer intelligiblen moralischen Welt tendierende Position scheint aber weder sachlich überzeugend noch die Position Kants zu sein.

Kant zufolge gründet die Ethik nicht darin, dass ich intuitiv eine real existierende, intelligible moralische Welt einsehe. Sein Grundlegungsverfahren hat vielmehr folgende zwei Schritte. Zunächst zeigt er mithilfe einer analytischen Methode, und dabei einmal im Ausgang von der alltäglichen moralischen Urteilspraxis, dass das einzig mögliche Moralgesetz das formale Gesetz der Gesetzmäßigkeit der Maximen ist (Kant 1968a, pp. 393–421; Kant 1968b, pp. 19–30); dieses Gesetz könnte jedoch eine bloße chimärische Idee sein, ohne mögliche Anwendung und ohne tatsächliche Geltung. Um diesem Vorbehalt zu begegnen, leitet Kant in der Lehre vom Faktum der Vernunft (Kant 1968b, pp. 42–50) die praktische Realität, das heißt die reale Möglichkeit der Gesetzesgeltung, von seiner Wirklichkeit ab: Weil die reine praktische Vernunft in einer lateinisch *factum* genannten freien Tat *wirklich* meine Willkür bestimmt, erzeugt sie durch diese Bestimmung das hervorgebrachte Faktum der Geltung des Gesetzes, das sich in meinem Bewusstsein der Verbindlichkeit des Gesetzes bekundet. Worauf es hier in unserem Zusammenhang ankommt, ist, dass die tatsächliche Geltung des Gesetzes von dem *performativen Vollzug* abhängt, in dem die fungierende reine praktische Vernunft in mir meine Willkür bestimmt, und dass das *unmittelbare Bewusstsein* dieser Gesetzesgeltung das *Ergebnis* dieser performativen Bestimmungsleistung ist. Die ‚reale‘ Geltung des Gesetzes und das ‚intuitive‘ Bewusstsein derselben gründen damit nicht in ontologischen Tatsachen einer intuitiv fassbaren moralischen Welt, sondern in jenem Vollzug, der innerhalb eines moralischen Subjekts selbst performativ stattfindet und sich infolgedessen aus der Perspektive der Ersten Person Singular im Bewusstsein bekundet. Dass der Mensch Zweck an sich ist, wird von Kant hingegen systematisch erst in der Einleitung zur Tugendlehre in der *Metaphysik der Sitten* im Rahmen der Deduktion des obersten Prinzips der Tugendlehre abgeleitet (Kant 1968c, pp. 394–395). Sein Argument ist dort *in nuce*, dass die reine praktische Vernunft im Menschen eigene, moralische Zwecke setzen muss, um den Zwecken der Neigungen etwas entgegenhalten zu können, und sie sich als Quelle dieser moralischen Zwecke notwendig selbst zum absoluten Zweck setzen muss, wenn es sie widerspruchsfrei als Zwecksetzungsvermögen geben soll. Auch die Zweckansichthaftigkeit ist damit keine ontologische Eigenschaft einer vorhandenen Entität, die intuitiv erfassbar wäre, sondern sie hängt von dem *performativen Vollzug* der moralische Zwecke setzenden reinen praktischen Vernunft ab.



Kant scheint nun durchaus in einem weiten Sinne ‚phänomenologisch‘ vorzugehen, wenn er im ersten Abschnitt der *Grundlegung* den Gesetzesinhalt analytisch aus den spontanen Urteilen der moralischen Alltagspraxis herausfiltert und in der zweiten Kritik die Möglichkeit seiner Geltung über die Wirklichkeit des Bewusstseins dieser Geltung sowie deren Beleg für eine freie Tat der legislativ fungierenden reinen praktischen Vernunft in uns auffasst. Weshalb sollten wir nicht einfach hier stehen bleiben und uns zu Kantianern erklären? Es gibt insbesondere drei Gründe, die es aus der Perspektive einer Phänomenologie des Ethischen nötig machen könnten, doch über Kant hinauszugehen. In Bezug auf alle drei lässt sich an Emmanuel Levinas anknüpfen. Erstens fragt Levinas tiefer in die Quelle jener moralisch-praktischen Vernunft hinein. Während Kant schlichtweg davon ausgeht, dass es in uns eine fungierende reine praktische Vernunft gibt, führt Levinas die Stiftung jener nicht instrumentellen Vernunft in seinem Spätwerk *Autrement qu'être ou au-delà de l'essence* (Levinas 1978) darauf zurück, dass der Andere mich immer schon angesprochen, zur Verantwortung gerufen und derart recht eigentlich erst mein Selbst konstituiert hat. Der performative Vollzug einer moralisch-praktischen Vernunft wird damit auf das immer schon erfolgte intersubjektive Anspruchsgeschehen zurückgeführt. Nur weil der Andere mich immer schon angesprochen hat, kann ich überhaupt dazu veranlasst werden, nach Handlungsgründen zu suchen, die mehr als nur mich selbst betreffen. Es gibt nicht einfach reine praktische Vernunft in uns sowie es einen Herzschlag gibt, sondern es gibt sie, weil der Andere mich immer schon so angesprochen hat, dass ich mich dazu verbunden fühle, nach Handlungsgründen zu suchen, die ich auch ihm gegenüber rechtfertigen zu können meine. Zweitens scheint die derart gestiftete reine praktische Vernunft nicht immer schon, wie Kant meint, auf die Frage nach der möglichen Gesetzmäßigkeit meiner Maximen für *alle* vernünftigen Wesen bezogen zu sein. Im Anspruch des Anderen spricht mich zwar der Dritte stets mit an, ein Anspruch, in dem durchaus die kantische Aufforderung zur Bildung von Maximen gesehen werden kann, von denen ich wollen können muss, dass sie allgemeine Gesetze für jedermann werden. Der ursprüngliche Anspruch des Anderen aber stiftet eine ethische Bindung, die noch diesseits des Dritten und damit noch diesseits der *Frage* nach allgemeiner Gesetzmäßigkeit für *alle* vernünftigen Wesen liegt. In dem Anspruch der von ihm so genannten „*an-archische[n] Vernunft*“ (Levinas 1998, p. 361 [1978, p. 212]) denkt Levinas eine Vernunft, die jenseits von Kants *bloß* praktischer, instrumenteller Vernunft liegt, aber diesseits von Kants immer schon auf Gesetzmäßigkeit für *alle* vernünftigen Wesen bezogenen reinen praktischen Vernunft. Die kantische Geltung eines Gesetzes reiner praktischer Vernunft wird bei Levinas in einer durchaus auf Kant zurückführbaren Tradition zur Stiftung eines Leitfadens, der der kreativen Beantwortung singulärer neuer Ansprüche zugrunde liegt, aber durch keine konkrete Regel jemals er-

schöpft werden kann, weil er als Vernunftidee das Moment eines praktisch Unendlichen impliziert. Diese Neuerung gegenüber Kant bei Levinas entspringt dem Umstand, dass bei Levinas die Herausforderung der Intersubjektivität und wesentlichen Pluralität ins Zentrum rückt, welche für Kant noch nicht eigens zum Problem geworden war. Das hat drittens zur Folge, dass es bei Kant zumindest unterbelichtet bleibt, weshalb ich nicht nur mich selbst, sondern auch die Anderen als Zwecke an sich selbst auffassen muss. Auch auf diese Frage aber scheint sich im Ausgang von Levinas eine Antwort finden zu lassen. Wenn die moralisch-praktische Vernunft allererst durch den Anspruch des Anderen in mir gestiftet wird, dann muss ich nicht nur die reine praktische Vernunft in mir, sondern auch die Quelle ihrer Stiftung im Anderen als Zweck an sich selbst auffassen. Der Andere ist Zweck an sich selbst, weil er die Quelle der Stiftung reiner praktischer Vernunft in meinem Selbst ist, die immer schon geschehen ist und mich allererst zu einem Zweck an sich selbst macht.

Eine derartige, hier nur überaus skizzenhaft durch drei Perspektiven ange deutete, von Levinas inspirierte kantianische Phänomenologie des Ethischen könnte sowohl hinsichtlich der in ihr implizierten Deutung der kantischen Ethik als auch systematisch überzeugender sein als ein kantianischer Intuitionismus der oben erörterten Art. Kant selbst ist der Auffassung, dass es die Dimension des Ethischen überhaupt nur auf der Basis eines *performativen Vollzugs* gibt, in dem eine fungierende reine praktische Vernunft die Willkür bestimmt. Verschwindet dieser Vollzug, dann gibt es auch kein geltendes Gesetz und keine Dimension des Ethischen mehr. Die Situation eines derartigen Verschwindens wäre nicht eine solche, in der eine an sich bestehende Welt moralischer Tatsachen uns bloß *intuitiv unzugänglich* würde, sondern es *gäbe* das Ethische schlichtweg nicht mehr. Sowohl Kant als auch Levinas sind dieser Meinung, wobei Kant allerdings die reine praktische Vernunft noch für ein Konstitutivum des Menschen hält, während Levinas in dieser Hinsicht deutlich skeptischer ist.<sup>3</sup> Kant und Levinas müssen damit keine real existierende moralische Welt annehmen, die wir im Idealfall intuitiv erfassen. Sie scheinen aber nichtsdestotrotz die Vorzüge des ethischen Intuitionismus nicht zu verspielen. Die Intuitionisten suchen einerseits der Pluralität ethischer Erfahrung Rechnung zu tragen und andererseits nichtsdestotrotz einen ethischen Relativismus zu vermeiden. Da ein von Kant und Levinas her kommender Ansatz den Inhalt des moralisch Guten und Gebotenen völlig offen lässt und nur der Dimension formaler Vernünftigkeit unterwirft, muss er die

---

<sup>3</sup> Levinas zufolge leben wir heute in einer „göttlichen Komödie“, einer „Zweideutigkeit von Tempel und Theater“, in der unsere Existenz durch ein grundlegendes Schillern zwischen der Dimension des Ethischen und des Nichtethischen geprägt ist, in der das Ethische schlechthin zu verschwinden droht (Levinas 1981, p. 108 [2004, p. 115]).

Vielfalt und den Wandel der Lebensverhältnisse sowie den Unterschied zwischen den Menschen nicht im Vorhinein unter ein inhaltlich bestimmtes Prinzip des Guten subsumieren. Nichtsdestotrotz vermag er einen ethischen Relativismus zu vermeiden. Zwar sind die Verbindlichkeiten relativ auf eine fungierende reine praktische Vernunft, und diese kann es auch nicht geben; wenn sie mit ihrem einzig möglichen Gesetz im intersubjektiven Anspruchsgeschehen aber einmal gestiftet ist, dann haben die aus ihrem Zusammenspiel mit der lebensweltlichen Erfahrung entspringenden Verbindlichkeiten die ganze Stärke, die sie benötigen, um uns zu binden, denn sie sind dann keineswegs nur vermeintliche Verbindlichkeiten, die manche Menschen in manchen Zeiten *de facto* für gültig halten.

Gibt es also einen kantianischen Intuitionismus? Man könnte in Bezug auf unsere Überlegungen zwar davon sprechen, dass wir ‚intuitiv‘ durch den erfahrenen ethischen Anspruch die ‚Realität‘ des performativen Vollzugs der fungierenden reinen praktischen Vernunft und der Zweckansichthafteit ihres Trägers erfassen; dieser Sprachgebrauch jedoch mutet irreführend an, weil die Begriffe ‚Intuition‘ und ‚Realität‘ in einem ganz anderen Sinne verwendet würden als im klassischen Intuitionismus, wo sie Intuitionen meinen, die sich auf objektive moralische Tatsachen richten. Im Rahmen einer kantianischen Phänomenologie des Ethischen erscheint es daher sinnvoller, auf diese Termini zu verzichten und von einem ursprünglichen ethischen Anspruch zu sprechen, der sowohl eine performativ fungierende reine praktische Vernunft als auch die daran geknüpfte Zweckansichthafteit des Selbst und des im Anspruch erfahrenen Anderen bezeugt. Die zahlreichen Verbindlichkeiten aber, die dem Zusammenspiel einer einmal gestifteten reinen praktischen Vernunft mit der wechselnden Vielfalt der Ansprüche und Lebensverhältnisse entspringen, sind so stark und drängen sich erfahrungsmäßig so direkt auf, dass man meint, es mit Intuitionen einer an sich existierenden Welt moralischer Tatsachen zu tun zu haben.

## Bibliographie

- Audi, Robert (2004): *The Good in the Right. A Theory of Intuition and Intrinsic Value*. Princeton/Oxford: Princeton University Press.
- Heinrichs, Bert (2013): *Moralische Intuition und ethische Rechtfertigung. Eine Untersuchung zum ethischen Intuitionismus*. Münster: Mentis.
- Kant, Immanuel (1968a): *Kritik der reinen Vernunft (1. Auflage 1781) – Prolegomena – Grundlegung zur Metaphysik der Sitten – Metaphysische Anfangsgründe der Naturwissenschaft*. Berlin/New York: de Gruyter (= Akademie Textausgabe. Bd. IV).
- Kant, Immanuel (1968b): *Kritik der praktischen Vernunft – Kritik der Urtheilskraft*. Berlin/New York: de Gruyter (= Akademie Textausgabe. Bd. V).

- Kant, Immanuel (1968c): *Die Religion innerhalb der Grenzen der bloßen Vernunft – Die Metaphysik der Sitten*. Berlin/New York: de Gruyter (= Akademie Textausgabe. Bd. VI).
- Levinas, Emmanuel (1978): *Autrement qu'être ou au-delà de l'essence*. Den Haag: Martinus Nijhoff [dt. *Jenseits des Seins oder anders als Sein geschieht*, übersetzt von Thomas Wiemer. Freiburg/München: Alber <sup>2</sup>1998].
- Levinas, Emmanuel (2004): *De Dieu qui vient à l'idée*. Paris: Vrin, 2. Auflage. [Dt. Teilübersetzung: Abschnitt „Gott und die Philosophie“. In: Bernhard Casper (Hg.): *Gott nennen Phänomenologische Zugänge*. Freiburg/München: Alber 1981, pp. 81–123.]
- Rawls, John (2000): *Lectures on the History of Moral Philosophy*, hg. von Barbara Herman. Cambridge (Massachusetts)/London (England): Harvard University Press.
- Ross, William David (2009): *The Right and the Good*, hg. von Philip Stratton-Lake. Oxford: Clarendon Press.
- Schönecker, Dieter/Wood, Allen (2007): *Kants „Grundlegung zur Metaphysik der Sitten“*. Ein einführender Kommentar. Paderborn: Schöningh, 3. Auflage.
- Stern, Robert (2012): *Understanding Moral Obligation. Kant, Hegel, Kierkegaard*. New York/Cambridge: Cambridge University Press.
- Schönecker, Dieter (2013): „Das gefühlte Faktum der Vernunft“. In: *Deutsche Zeitschrift für Philosophie* 61. No. 1, pp. 91–107.



Roberta De Monticelli

# Sensibility and Values Toward a Phenomenological Theory of the Emotional Life

**Abstract:** Affects – quite especially, but not exclusively, bodily affects – “shape the mind” by general recognition – and emotions are more and more popular as a topic of philosophical research. Since Antonio Damasio’s well-received work proposing the “embodied mind” approach to emotions, feelings have been investigated with brilliant results from both a phenomenological and a neurobiological perspective (for recent examples see Ghallagher-Bower 2013 and Fuchs and Koch 2014).

On the other hand, philosophers have become aware of the fundamental importance of the realm of feeling in both the cognitive and practical exercise of reason. Yet, although the relation between emotions and values is much discussed in analytic philosophy, contemporary research seems to be lacking a general theory of feeling, which would somehow connect the two levels of affective sensibility apparently concerned: one that is basically embodied, and one that is cognitively of a “higher” level, involved in a large variety of acts and behaviors characteristic of a rational and moral agent – such as a human being.

This paper presents an outline of such a general theory. It draws on classic sources in phenomenological literature, yet aims to provide a somewhat independent response to some of the main questions in contemporary debates, including the crucial one concerning the objectivity/subjectivity of values and value judgments.

## 1 Recent Approaches to Emotional Life. What Is Missing

In his influential book on emotions, Peter Goldie – without being very explicit about it – seems to use this word in at least two ways: in a very broad and comprehensive way – synonymous with “emotional feeling” (Goldie 2000, p. 51), including bodily feelings and moods (pretty much like the very comprehensive sense of “passions” in Descartes, or in the classical philosophical tradition) – and in a narrower way, *emotions* as contrasted to moods and bodily feelings. That there is a very broad philosophical sense in which Goldie uses the word

“emotion” is also proven by the general criterion he provides to distinguish emotions from “brute feelings like toothache, which we cannot make sense of” (Goldie 2000, p. 20): meaningfulness, or intelligibility. This criterion is, by phenomenological standards, both acceptable intensionally and too restrictive extensionally: as a way of becoming aware of part of one’s body, or of one’s fragility or poor health, even a toothache can make sense, let alone a physical pleasure, a bodily feeling, a mood.

One of the purposes of the theory of feeling outlined here is actually to exhibit a rationale for distinguishing, ordering and connecting different phenomena of feeling that most classic and contemporary philosophical literature piles up in a heap or a list where one can find toothaches or the pleasures of the table, along with the sin of pride or the passion for truth – giving rise to what I will call the hodgepodge problem. For we must find criteria for *ordering* a taxonomy which will allow us to include phenomena of such different classes as *feelings of the senses* (pleasure, pain), *bodily feelings and moods* (being tired, disgusted, being anxious), *emotions of different kinds and levels* (basic: fear, anger; non-basic: shame, regret, guilt, indignation); *personal feelings or sentiments* (love and hate, esteem, respect or contempt), *passions* (jealousy, gambling addiction, passion for truth), *habitual attitudes of personality* (self-esteem, confidence, humility, curiosity...).

## 2 Goldie’s approach: Some unanswered questions

Goldie’s (2000) book in a sense seems to reject a systematic ambition:

I do not put forward a single, central claim and then seek to defend it against opposing positions [...]. [This book] takes as the phenomena not just the emotions, but looks more widely to related phenomena such as consciousness, thought, feeling, imagination, interpretation of action out of emotion and of expression of emotion, moods, and traits of character; emotion cannot be considered in isolation from these other topics. (Goldie 2000, p. 1)

Yet Goldie presents us with at least four claims that any phenomenologist would share.

- (1) An adequate account of emotions cannot be given from the impersonal stance of the physical sciences;
- (2) Emotions are reasons for action;
- (3) Emotions have intentionality;
- (4) Emotions should not be “over-intellectualized”.

Let's briefly scrutinize these claims. As for (1), Goldie is very clear in opposing the personal and impersonal stance, by being careful not to equate the term "impersonal" with "third-personal": "As a quite general idea, when I think or talk about other people, I can do so third personally, without losing sight of the fact that these other people have a point of view." (Goldie 2000, p. 2). Indeed, the fact that someone else has a point of view, just as I do, is presupposed in this way of talking. In this sense we do adopt a personal stance even when speaking of third persons. Goldie's claim is that an account of emotions that adopts "the impersonal stance of the physical sciences" (Goldie 2000, p. 2) would explain away its very object. Now one may ask: why? There are several answers to this question. One of them would be that even if emotional intentionality can be naturalized, the associated "phenomenal consciousness", the subjective state or *qualie* accompanying the intentional state, *cannot*. Yet Goldie could not subscribe to this answer, given his claim (4), as we shall see. Another available answer is the phenomenological one I shall try to outline at the end of this paper (cf. section 3).

As for (2), Emotions are reasons for action: a most interesting thesis, given Goldie's remark:

"[...] it is a familiar enough notion in philosophy that rationality is essentially *normative*" (Goldie 2000, p. 2). If we agree that this familiar notion represents a true statement, emotions provide a clear example: they apparently are *reasons for action*; moreover, they can be *good* or *bad* reasons. Emotions are regarded as *appropriate* or not, *proportionate* or not. As Goldie remarks, we do distinguish between "having an emotion" and "being emotional" (Goldie 2000, p.48). Indeed we have countless ordinary language refutations of the old, but still pretty widespread prejudice about the irrationality of emotions. Just compare the effect of attributing a delicate sensibility to a person, with that of reproaching her for her rash emotionality. But philosophy should give our judgements a better foundation than common language. So, how are we to justify our attributions or denials of appropriateness or proportionality to emotions and consequent actions? In a sense, this is *the* question a philosopher should answer, because of its fundamental implications for ethics and practical philosophy generally.

An answer to this question from Goldie's point of view, nevertheless, seems to depend on Goldie's treatment of the intentionality of emotions. For aren't appropriateness and proportionality somehow linked to the *adequacy* of emotional responses to actual objects or states of affairs?

Then, concerning claim (3), the question is: if emotions have intentionality, exactly what is their intentional object – or the "formal object" of emotion? For if there is no general intentional/formal object, then, first of all, how can we respond to the question of normativity, and, secondly, what would be the point of claim (4)?



Claim (4), in fact, is way less harmless than it seems, it is indeed a real challenge thrown down to the entire tradition of “cognitive” (or, better yet, “cognitive-conative”) accounts of emotion.<sup>1</sup> The reason emotions are currently “over-intellectualized” is precisely because of the burden of intentionality, which usually operates on the cognitive component of the content emotions are analyzed into – this cognitive component generally being a belief. We may illustrate this procedure with classical Aristotelian or Davidsonian accounts, according to which, for example, hope can be analysed into a belief that it is possible that *p* will be the case combined with the desire that *p* be the case. This seems to leave no intentional role for the feeling component of hope, which must be smuggled in, so to speak, in the form of the “phenomenal consciousness” or qualitative component of hope – that is, as a state of mind “added on” to the intentional state. This is why Goldie takes a position against “*the add-on view*”, by introducing the notion of “feeling towards”:

So, my position can be seen as retaining what is right about the traditional view that intentionality is essential to emotion, but bringing in feeling in the right place, as an ineliminable part of the intentionality of emotional experience. (Goldie 2000, p. 4)

A phenomenologist could not but agree. And yet: *how* does *feeling* manage to have intentionality? As soon as we look for further analysis of this point, we find it, but it is rather disappointing: “Feeling towards is *thinking of* with feeling, so that your emotional feelings are directed towards the objects of your thought” (Goldie 2000, p. 19). Here, the degree of over-intellectualization does not seem so dramatically reduced with respect to the belief-desire account!

This entire discussion can lead to a basic question: is there *a theory of feeling* underlying Goldie’s claims, and supporting them? If a theory has to be forced out of the *coherent* answers that Goldie could give to the questions raised above, then there seems to be no systematic theory underlying Goldie’s claims. I shall try to put forward some principles of a systematic theory, which will allow us to answer these questions.

---

<sup>1</sup> Cf. Aristotle (1975), Davidson (1980), De Sousa (1987).

### 3 The phenomenological approach: Some principles toward a theory of feeling

Let's first state three basic claims or principles that constitute the very foundation of a phenomenologically acceptable theory of feeling:

1. Feeling is essentially a perception of the value-qualities, whether positive or negative, of things.
2. Affective experience in all its parts (including its conative aspects, drives, desires etc.) is founded on emotional sensibility.
3. Sensibility has a structure of layers ("stratification"), corresponding to an objective hierarchy of value-spheres.

These principles suggest some answers to the questions raised above. In particular, 1 and 3 should serve to shed light, respectively, on *the objective and subjective poles* of the intentional relation – thus explaining what is special in that mode of consciousness which comprises *feeling* consciousness, both in the way the *objects of feeling* present themselves to the subject and as a mode of presence of *the personal subject to him or herself*.

In other words, the principles should serve to shed light on the two poles of emotional experience *as experience*, that is, as a specific mode of direct cognition or acquaintance with reality, which, as a form of perceptual cognition, can be veridical or illusory (so that we can correct illusions and "learn from experience").

Such a general account of the two poles of emotional intentionality, parallel to the classic accounts Husserl or Merleau-Ponty provide for perceptual intentionality, is indeed required if by experience we mean a mode of *cognition*, i.e. of consciousness, which, unlike pure fantasy, is subject to normativity of some kind (emotions are right or wrong in some specific sense).

Claims 1 and 3 actually provide a basis for answering questions about the rationality of emotions as good or bad reasons for action: their appropriateness, their proportionality, or the lack of one or both of these features. Claim 2 should contribute to the explanation of the specific nature of the intentional relation itself, thereby allowing for a more thorough analysis of the complexity of what Goldie calls "episodes of emotional experience, including perceptions, thoughts and feelings of various kinds [...] dispositions [...]" (Goldie 2000, p. 13), further experiences, behaviours. This can be seen in the example Goldie provides from Tolstoy's *War and Peace*: Pierre's sudden realization that he is in love with Natasha ("at that moment Pierre involuntarily betrayed to her, to Princess Mary, and above all to himself, a secret of which he himself had been unaware": Goldie 2000, p. 15), and the explosion of inner and outer life unleashed by this

discovery: joyful flushing, painful distress, confused speech, new perceptions, new questions, new thoughts and behaviours ... Claim 2, taken together with claim 3, also provides an argument for the first of Goldie's unsupported claims, on the impossibility of eliminating a personal stance from an account of any bit of the emotional life of a human being.

### 3.1 Feeling as perception of value-qualities

My first claim concerns the specific character of feeling's intentionality in relation to its object: feeling is the *mode of presence* of the value-qualities of things (salience, "affordances" in J.J. Gibson's terminology<sup>2</sup>, tertiary qualities of all sorts). This principle addresses the question of the formal object of emotional intentionality. It points to a philosophically widely neglected, yet pervasive feature of the life world: the plurality, richness and variety of positive or negative value qualities "colouring" things, events, states of affairs, situations in the surrounding world. Indeed it is hard to find qualifying words in our languages, adjectives, which do not refer to some value quality. Axiology is in a sense the ontology of adjectival language.

So, the first piece of evidence for this principle is both semantic and phenomenological. We actually are presented with an extremely rich variety of apparent value-qualities of things by means of *feeling*. I feel the unpleasantness of a sting, the bodily or psychological discomfort associated with a state of illness or weariness, the agreeable nature of an arrangement of colors. But I also sense the nobility of a gesture, the vulgarity of an attitude, the wickedness of an act, the beauty of a masterpiece. The harmonious way a tool or a piece of furniture fits one's body, the peaceful form of a teapot are among the "affordances" of an object. These qualities too are somehow "perceived": feeling is the appropriate mode for this kind of perception. Emotion is in this respect essentially *like* perception: it is, as the German language has it, a *Wertnehmen*. The functional and aesthetic qualities of artifacts are not only "seen": they are felt. This feeling is always accompanied by the exercise of other functions, both sensory and otherwise. From a phenomenological perspective, different sorts of states— sensory perceptions, memories, experiences of empathy or social cognition, understood

---

<sup>2</sup> "I have coined this word as a substitute for *values*, a term which carries an old burden of philosophical meaning. I mean simply what things furnish, for good or ill. What they *afford* the observer, after all, depends on their properties." (Gibson [1966] quoted in Jones 2003, p. 111)

speech acts, etc. – *motivate* a feeling, and further acts—new perceptions, speech acts, choices, actions etc. – are in turn *motivated by* such a feeling.

A second piece of evidence is provided by remarks such as the following, which Goldie makes almost incidentally in reference to the education of emotions. An “evaluative property” such as “being dangerous” does not correspond to a general, shared feature of objects and hence is no “natural” property. Yet in our infancies we are taught to recognize danger and to respond to it correctly – namely with fear – or to give in to fear only where appropriate. “The essential idea is that our emotions can be educated.” (Goldie 2002, p. 28)

Such remarks immediately give rise to the following question: is there anything *veridical/illusory in the feeling itself which reveals that which is dangerous*, in spite of its being a “non-natural” property? In other words: is this “evaluative property” a quality things actually have or fail to have?

If it is a quality things actually possess then feeling is an – albeit partial and fallible – way to access reality, under the aspect of value. Fear itself, or rather the sensitive core of it, is the way in which dangerous things reveal themselves as dangerous – even if further, more accurate experience of the thing concerned can correct a possible illusion and dispel fear. Then the intentionality of feeling is of a cognitive sort, its direction of fit is mind to world – and its formal object is value, as phenomenologists claim.

But if this is *not* the case, and the quality of being dangerous itself, being a “cultural” property, is just a culturally relative, subjective or conventional posit (as the made-up story of the anthropologist learning to identify behaviors motivated by recognition of a tribal value-property inaccessible to him, referring to the imaginary concept of “*gopa*”, tends to suggest),<sup>3</sup> then what does it mean that “an evaluative property” is “a property whose recognition *merits* a certain sort of response,” and that when facing something which is dangerous “the response which is merited is one of *fear*” (Goldie 2002, p. 34)? What can the language of appropriateness and proportionality convey, if not a sort of adequacy to *what is required* by the phenomenal thing itself? And if we give up the idea of some sort of “requiredness” (Goldie 2002, p. 31) of an objective kind, how can we maintain that there is a rationality of emotions? Any “emotivist” or “expressivist” view of emotion is perfectly compatible with a purely conventionalist view of norms, appropriateness, proportionality.

Let’s summarize our claim: Far from being *opaque* and *irrational*, emotional experience is essentially the mode of presence of the qualities of value of things and is also the more or less appropriate response, and therefore the more or less

---

<sup>3</sup> Goldie 2002, p. 31.

correct, proportionate, or “rational” response to the *requirements* that these present. We may call this the *first phenomenological principle of emotional life*. Feeling is the mode of presence of the most diverse axiological qualities, or values. This openness to axiological qualities makes up the component of *intentionality* characteristic of all the experiences of the emotional sphere. Thus, for the phenomenologist, this is a dimension of our experience of the real, by no means whatsoever a realm of subjective arbitrariness – nor is it a mere system of functional alarms or incentives necessary for the survival of the organism.

Openness to reality is synonymous with *fallibility*. *Biases, one-sidedness, illusions* of affective feeling, of course, are just as if not more frequent than those of sensory feeling – and therefore equally open to correction and “proof to the contrary”, that is to confirmation of a judgment contrary to the original one, if there was one. Here again, as in the case of the sensory domain, reflection may help to correct (a judgment based on) an erroneous perception: but it is in fact a new perception that will tentatively confirm the correction.

On the other hand, in value matters, poor or inadequate feeling, sentimental stupidity, short-sightedness or obtuseness are at the root of inappropriate axiological responses – and among these, of *ethically* inappropriate responses – to what the world and other people require or merit from us. The first phenomenological principle implies paying profound attention to the problem of the *education of feeling*, or emotional culture. This education of feeling can even be said to be at the heart of culture from the phenomenological point of view.

### 3.2 Sensibility as a foundation of emotional life

Let's come to our second claim: affective experience is founded on *sensibility*. This principle adds to the first one, which characterizes the essence of feeling – *the role of feeling in the complexity of emotional life*. In one way or another, feeling represents the basis of emotional life, but emotional life is certainly not reducible to feeling alone – feeling serves as the primary, non-reducible mode of *reception*. Sensibility serves as the foundation for the rest of emotional life, in the precise sense that it *motivates everything in this life which is not mere reception, but includes response and spontaneity*. In the end, feeling motivates volition. But first of all feeling forms the core of all sorts of emotions (in the proper sense) and passions.

Thereby we place the emphasis on *receptivity* as the fundamental aspect of emotion (in the broad sense). Contrary to most classic and contemporary approaches, the conative aspect (drives, desires) is founded, not founding. Let's clarify the point of emphasizing receptivity.

In fact, two “moments” are apparent in most emotional phenomena: (a) *Being affected by* or receptivity, “passivity”, being “struck” or “impressed” by something: in short, the *receptive component* of an emotional episode, a kind of perception. (b) *Being inclined to*, “moving” to or from, drives/desires, (*Streben*) or in short the *conative component*, the urge to action.

There have been two main trends in the analysis of emotional life. Either the intentionality and somehow cognitive character of emotions has been acknowledged, but in this case emotions have been, as Goldie would have it, “over-intellectualized”: that is, cognition has been attributed to a belief component (as in Aristotle, or Davidson). Or emotions have been thought of as irreducible to anything else, but then they have been typically characterized as “irrational”, “non-cognitive”, “purely subjective” or at most provided with a functional role for survival (vital utility – as in Descartes and Kant, or interpreted according to dualistic approaches, but also according to naturalistic views, such as Damasio’s).

In all these cases, what is thought of as essential or not eliminable is the conative component (drives, desires). Let’s consider the classical concept of *thumos*, which represents the emotional soul in Plato’s psychology, the word (at least as used by Homer) referring to any sort of passionate desire. Or consider the importance of both the appetitive and the aggressive components of the passions in Aristotle and the scholastic tradition. Even Augustine is incapable of conceiving of emotional life independently of the modes of desire, with the consequent restlessness, constitutive of our “created” nature. Franz Brentano echoes Aristotle; and then Sigmund Freud, echoing Arthur Schopenhauer, reproduces a Kantian dualism: nothing is more irrational than emotional life, rooted in sexual drive.

Against this tradition stands the second principle of a phenomenological theory of feeling. What evidence can we provide in its support? As Goldie once again would say: “being emotional” differs from “having an emotion”. Here we can go deeper into the meaning of this semantic remark. Foundation, in our use of the word, entails a relation of *ontological dependence*. That the receptive component provides the foundation for the conative one means, first of all, that the latter cannot be without the former, but the former can exist without the latter. There are lots of examples of this second possibility: aesthetic experience, such as listening to music and discriminating its aesthetic and expressive qualities; a mother lovingly contemplating her sleeping child; blissfulness; calm despair; surprise; amazement.

But what does the first relation described above amount to, exactly? What does it mean that the conative component cannot exist without the receptive one? Don’t we speak of hardness of heart, or of a person indifferent to other people’s sufferings. Feeling, even when obtuse or atrophic, even when conspicuous in its absence, is only relatively absent, *absent at the “required” level*: it simply

cannot be completely or absolutely absent. To see this point we need the further notion of layers of sensibility, which will be introduced as part of the domain of our third principle. An example will suffice for the moment. A radically egotistical person does not lack whatever feeling could motivate merciless behavior – for example the ability to sense feelings of pleasure and pain, vital feelings of comfort and discomfort – he/she may in fact prefer his/her peace of mind to going through the trouble of saving a drowning child. More generally, feeling, even when it is “missing” (to some extent), always seems to be a key determinant motivating emotional response, decision and action – and we will better see why (in section 2.3.3), since this proposition involves the notion of *preference*, i. e. of a (subjective, more or less adequate) ordering of values.

A paradoxical example of this is the inverse proportion often observed between sensitivity and excitability, or between the appropriateness of feeling and *emotional agitation*. Countless examples can be found in everyday life. Chekhov’s early work, the little literary gem *Ivanov*, presents a sort of theorem of this inverse proportion.

### 3.3 Structure and layers of sensibility

In order to clarify our third claim we must first consider a further dimension of feeling’s intentionality. Thus far we have considered one: the *breadth* of its domain, or horizon. The qualities of value – positive and negative to varying degrees – are *many*. The extraordinary *richness* of the negative or positive value qualities to which we are sensitive provides a basic starting point for any phenomenological reflection on feeling. There is virtually no situation in life in which at least some of these qualities are not currently or habitually present. I walk along the road, and the atmospheric conditions present me with an environment which is pleasurable to a greater or lesser extent; the urban architecture gives me aesthetic pleasure or pain – or something in between – with every step; the beggar on the corner draws my attention to the horrors of poverty; in words and images the newsstands scream violence and beauty, ferocity and injustice; every event along my way, whether serious or trivial, manifests qualities of value from some point of view: aesthetic, ethical, legal, economic, ecological, ergonomic, hygienic, gastronomic... Viewed *a parte subiecti*, this axiological richness of the world represents the quantity of values that affect us, touch us and move us, physically or emotionally. But also immediately given is the *quality of these positives and negatives – the way in which they touch us*: more or less deeply or intimately.

The other dimension which immediately reveals itself to phenomenological observation is therefore that of how *deeply* we are affected when we feel a qual-

ity of value, positive or negative – the ugliness of a pair of shoes on a friend’s feet, or of one of his gestures towards us. We also express this difference in speaking of the difference in *importance* or *weight* that things have. We are usually willing to distinguish the importance or weight that things have in themselves from their importance or weight for us under specific circumstances. For example, if it is important to me that my friend make a good impression on someone in a position to help him, his ugly shoes may disturb me more than a verbal insult he may direct at me – even if I continue to think that how one dresses is less important than kind behavior.

The difference in “objective” importance which we all recognize both in general and in many specific cases (killing is more serious than insulting; defacing a Beato Angelico worse than scrawling graffiti on a supermarket wall; giving one’s time or one’s life for a friend is more beautiful than giving her money for medical treatment) may be called *the rank of a value*.

Is this language of depth and rank more than metaphorical? Our third claim maintains that it is, that it does refer to an amazing kind of correspondence between the organization of our emotional experience and the structures of axiological meaning that any actual occurrence of goods and evils of some sort in a given situation exemplifies. Let’s recall our principle:

(3) Sensibility has a structure of layers (“stratification”), corresponding (or claiming to correspond) to an objective hierarchy of value-spheres.

Let’s first try to spell out its content more precisely, starting from the second term of the alleged correspondence: *an objective hierarchy of value-spheres*.

### 3.3.1 An objective ordering of value spheres?

Is there any such hierarchy? That there is one is of course the most controversial premise of our theory, this premise being subject to objections raised by axiological relativisms and culturalisms of all sorts. Hence the importance of focusing on the content of this premise in order to see what it actually claims and what it does not. One useful approach involves the faithful description of a relevant feature of our value experience, as attempted above: all the positive value-qualities we feel are present in things and that make them things that we consider good, or that are lacking and therefore “required”, and all the negative value-qualities we feel that make as many evils of the things they afflict – *these are not felt as having the same rank*. They seem to belong to *very different value spheres*, and *these value spheres are somehow ordered in a hierarchy*. So, what does this premise say exactly?



(I) It does not say that there is a complete, objective ordering of values, but of their *types* – of value spheres.

(II) It does not say that such a hierarchy is universally acknowledged or accepted, but only that:

- a) Such a hierarchy re-emerges throughout human history and during its most dramatic conflicts in the form of an *ordering of (spheres of) intrinsic value* (making a good an end in itself) which (spheres of) instrumental values are subservient to. From ancient Greek civilization to the most intimate core of religious experience of any faith, up to the tragedies of the contemporary world, this ordering places the values of personal flourishing over those of the prepersonal spheres, namely the vital and the social spheres – thus the values of the prepersonal spheres should be subservient to those of personal flourishing. To the vital sphere belong the values of life-enhancement, instrumental to personal flourishing, such as the satisfaction of all basic needs, health, good living conditions; to the social sphere belong all the values embodied by the institutions and good practices of a society – welfare, security, instruction, good administration, well-functioning economic life, and technical improvements in all fields – without which most people can have no chance of personal flourishing. But all this, such is the implication, should be the means or provide the conditions for realizing those values which bestow meaning to (all) personal lives: justice in matters of law and rights, beauty in nature and art, knowledge and joy in the quest for knowledge in all domains of science and morality.
- b) Nor does our premise imply that axiological truths within each sphere are universally acknowledged or accepted: but only that there are axiological discoveries much as there are scientific discoveries. Take justice as an example. There has been remarkable progress in our understanding of it – in our theories of justice – from the ancient Greeks to, say, the age of human rights. Our understanding increases through dramatic historical experience.
- c) Once this knowledge is accessible, not complying with it becomes unjustifiable. Today, one should not be allowed to openly support slavery or sexual discrimination. Even if there still are countless cases of violation of the human rights, their perpetrators usually don't claim to be right in acting as they do – they would rather reject the fault on their victims.
- d) The objective hierarchy theory does allow for a distinction between the importance or weight that things have in themselves, and the importance or weight that they have for us under specific circumstances.

It's impossible to argue convincingly for this part of our theory within the limited space of this paper; it is however important to have an intuitive grasp of it in

order more precisely to understand the content of claim (3). Nevertheless, there are more clues to its plausibility despite objections from disenchanting mainstream or axiologically relativistic views.

We do distinguish levels of gravity or importance in value-laden facts. The gravity of being deprived of an ice cream feels inferior to that of being raped. The importance of one's health is felt to be greater than that of a variety of pleasurable habits, e.g. smoking. And it's quite understandable that one may disregard gastronomic pleasures and even health for the sake of scientific or philosophical research...

### 3.3.2 Why a “stratified” structure of sensibility?

Let's now address the first term of the alleged correspondence: the structure of layers (“stratification”) that sensibility is supposed to have. To which problem is this claim trying to respond? We already mentioned it: the hodgepodge problem. A satisfactory theory of feeling should avoid lumping all emotional phenomena together. Indeed, this claim is a kind of structural or Gestalt principle for emotional life, parallel to the analogous principle phenomenologists have put forward regarding perceptual experience (De Monticelli 2013, p. 118).

Our emotional life is not at all chaotic, nor is it an unstructured flow of “states”. Against flow theories, we can point to motivational chains, which are “rational” (or irrational, in any event not simply causal). Having a sentiment of friendship for you motivates my joy at your coming in and my desire to get up and hug you – whose satisfaction I can choose to postpone because I'm in the middle of giving a lecture. (This power to endorse or reject motivations, thereby making them efficacious or powerless, distinguishes the motivational series from a causal series). But against hodgepodge theories, we may put forward our principle of vertical structure, or Gestalt. Here, the clearest description is offered by Max Scheler:

There can be no doubt that the facts which are designated in such a finely differentiated language as English by “bliss,” “blissfulness” [Glückseligkeit], “being happy” [Glücklichsein], “serenity,” [Heiterkeit], “cheerfulness” [Fröhlichkeit], and feelings of “comfort” [Wohlgefühl], “pleasure,” and “agreeableness” [sinnliche Lust und Annehmlichkeit] are not simply similar types of emotional facts which differ only in terms of their intensities [...]. (Scheler 1973, p. 330)

All these feelings share a positive quality; but – Scheler says – they do not necessarily differ in intensity. They differ – as we would say in ordinary language – in depth. What do we mean by that?

Intuitively, we realize that a feeling can touch a person more or less “deeply”, depending on *the degree of personal involvement*. For instance, the pleasure of a good ice cream, given my particular personality, satisfies me much less than the joy of understanding Plato. No doubt this joy will have *a higher degree of motivational power* than the pleasure of an ice cream: it might indeed motivate my choice to study philosophy instead of something else, with serious consequences for the rest of my professional life.

These two notions help to explain the metaphor in conceptual terms. But is it possible to give, if not a metric for depth, at least a rationale for the alleged ordering of layers of sensibility concerned, respectively, by the pleasure of ice cream and the joy of reading Plato? What is the rationale for this stratification?

A last passage from Scheler offers a powerful suggestion:

It is, for example, impossible for one to be “blissful” over happenings *of the same axiological level* that are “disagreeable” to another; the differences in these feelings also seem somehow to require *different axiological states of affairs*. (Scheler 1973, p. 331; emphasis added)

Here is the explanation. The “depth” of a feeling is proportional to the importance of the values concerned. Feelings are modes of presence of values *at different levels*.

### 3.3.3 Sentiments and value-preferences

We will now conclude our analysis of the third principle of a phenomenological theory of feeling, which, needless to say, would benefit from further investigation and evidence. Nevertheless, even in this context, the principle manages to add quite a bit to our theory. First of all it reveals a new feature of value-experience: we do often *feel* some kind of ordering of values, before and independently of an act of propositional judgment: a value quality presents itself to us in the form of a perceived priority, which we (fallibly) feel to be an objective one. We feel that some things that are good are more valuable than others, even when we decide that in a particular situation we will *give* priority to the less valuable. This feeling of an objective priority can be as illusory as any other feeling but is equally open to correction through further experience. A value preference is an act of feeling (hence reception) before it can motivate volition. True, a *felt* priority – a structure of preference – is not always present (otherwise no decision would ever present a dilemma), and even when it is present, it requires a sort of endorsement on the subject’s part. This is typically the case with all the feelings which involve “a [small] background hierarchy”, and a consequent “yes” or “no”: for example,

loving this particular person means feeling these specific value qualities to be more important than certain others. The same is true of admiration, contempt, sympathy, religious devotion, political passions...

Those feelings which involve such consent (or dissent) might be called *sentiments*. We can already see from the above examples that our third principle allows us to introduce an important explanatory tool, which seems to be missing from most theories of emotions: the notion of a *sentiment* explains the difference between the “primary” emotions (that we share with many other animals) such as fear, anger, sexual drive, intraspecific affections or hostilities etc. – and “secondary” or more refined emotions, typically human. Let’s consider a simple example. Young children don’t seem to experience the emotion of shame, not even relative to the very different cultural norms of decency which exist in different places and times. It is only when they are mature enough to implicitly consent to a new “small value hierarchy” involving a personal disposition to act, perceive, judge in new ways – in short, when they acquire the *sentiment of shame* with modesty – that they become capable, given the right circumstances, of the experience of blushing with shame, that is of experiencing the appropriate *emotion* (in the narrow sense). The German language makes it clearer, in this case, by distinguishing *Schamgefühl* (as a sentiment, *pudeur* in French) from *Scham* (as a secondary emotion, *honte* in French).

Here are some further merits of this principle:

- a) it completes the description of feeling’s intentionality: the depth of a feeling should be proportional to the perceived importance of its value-level.
- b) thereby explaining the rationality of emotions: spelling out what it is for a feeling experience to be adequate, that is appropriate and not disproportionate.

We can easily see why despair over a cancelled flight (in most, yet not all, cases) is disproportionate; or why to be like Adolf Eichmann, the Nazi official whom Hannah Arendt chose as a paradigm of the “banality of evil”, is to have horribly inappropriate feeling responses to states of affairs of extreme moral gravity. In fact, the documents of the Jerusalem trial frequently report that he regarded as “disagreeable” (like a bad meal) the murder of thousands of innocents.

### Sensibility and Selfhood

There is one last question we raised earlier, that is still waiting for an answer. It concerns Goldie’s first claim, against the “naturalization” of feeling. Why is it that satisfactory accounts of emotions cannot be given from the impersonal stance of the physical sciences?

Any phenomenologist would answer: emotional phenomena are not independent of personhood and individual personality, because they are *constitutive* of both. They are individuating and individuated! Or, in the richness of their contents and implications, they cannot be separated from the experiencing person (as opposed to mathematical discoveries, or even philosophical *thoughts*, that, as Frege used to say, “have no master”). That’s why we seek to understand different types of ill-fated love, perhaps, by reading Anna Karenina or Madame Bovary (that is, in the background of a personal narrative, as Goldie would say) much more than by studying academic psychology.

Yet this might seem a kind of begging the question. Can’t we give a more precise explanation? We can. We said that our third principle was a sort of Gestalt principle for emotional life. We saw how a vertical structuring of felt preferences works to ground the appropriateness and proportionality of emotions (or lack thereof). But wherever there is a form of Gestalt as an organizing principle of some domain of pieces of data, we must *look for the whole which integrates those pieces of data* – a whole which is not reducible to a mere sum of its parts, if such a Gestalt is to hold (De Monticelli 2013, pp. 118–120).

Our question, then, becomes: What is the “whole” of which feelings are the parts or “moments”? Our answer is, not surprisingly: the Self. By “the Self” I mean a person, *as she experiences herself, or as she is “given to herself”*, in that quite peculiar, irreducible way, in which anybody is not given to anybody else – that is, from a first person perspective (Baker 2013, p. 55).

From this perspective, layers of sensibility *simply are layers of Self, as they are experienced from within*. This explains the relation, already referenced (section 2.3.2.), between the depth/importance of a layer of sensibility and the *degree of personal involvement*.

Life and science teach us the role feeling exerts in shaping personality. A person who is wounded in her feeling is a person wounded in her deepest self. Or think of poor Phineas Gage, the construction foreman from Vermont who suffered extensive cerebral lesions in a work accident, whose case Antonio Damasio described in his highly successful book *Descartes’ Error* (1994). After recovering from the accident, he “was no longer Gage”. For the mental (or) psychological functions (or the parts of the brain) affected by the accident were exactly those that are indispensable for the modulated exercise of feeling.

Does our theory provide us with criteria for *ordering* a taxonomy of feelings like the one presented in section 1? Well, it does at least suggest an intuitive direction of deeper and deeper self-awareness, or rather *experience* of oneself. From the surface of one’s body to one’s “intimate” self.

Not all feeling experiences, indeed, come to be felt from inside as “really” concerning oneself – one’s deepest self, so to speak. Our theory suggests that

*there is a specific layer of sensibility constituting individual personality.* Call it the layer of sentiments, or the value-preferences ordering layer. If our theory is plausible, this is the very central, or most intimate, layer of selfhood. Call it the core-self: it is what ordinary language still calls a person's "soul".

Arguing this last claim in detail would require a further article, if not a whole book. A kind of map of the areas to be explored is outlined below in this trial picture of sensibility and selfhood, as a theory of stratification would describe it – which will bring this paper to an (open) conclusion.

## **4 Self from Surface to Depth: a Trial Schema.**

This schema summarizes the proposed taxonomy of feeling experiences by exhibiting its ordering principle in the correspondence between the layer of sensibility concerned by each (type of) experience and the layer of (one) self that the same experience "reveals". The increasing role that each (type of) feeling plays in personal life (its motivating power, its relative degree of personal involvement) explains the deeper and deeper self-revealing power which that feeling experience exhibits. Examples of each class in this taxonomy of feelings should make the series of correspondences more intuitively clear.

**Table 1:** Self from Surface to Depth. The Stratification Theory of Sensibility and Selfhood: A Tentative Schema

Examples	Layers of Sensibility	Experienced Layers of Self	Role in Personal Life	Value Spheres
Enjoying a massage	Affective sensations	Parts of one body	Foundations of (prereflective) self consciousness	Sensory values (pleasant/painful) Affordances
Hunger, tiredness	Bodily feelings	One's body (globally felt)	Indicators of one's vital or existential needs	Vital values Life enhancing or impairing
Anxiety, depression	Moods	One's global state ( <i>How are you?</i> )		Socially implemented: Utility sphere → Welfare Instrumental to culture and civilization spheres
Scare; anger; Basic desires and drives	Basic emotions	One's sudden changes of global state + conative component	Motivating personal behaviour	
Esteem; <i>Schamgefühl</i>	Sentiments/habitual dispositions of sensibility	One's dispositions involving value preferences ( <i>Who are you?</i> )	Indicators of personality	Values defining Cultures and their normative structures: Justice → Public institutions Beauty → Art and culture (Knowledge of) Truth → Science, education
Admiration; <i>Scham</i>	Higher emotions	Actual life of these dispositions, possibly involving conations	Motivating personal behaviour	Manifesting personal axiological orders of priority and preference
<i>Ressentiment</i> ; Passion for truth	Passions	Sentiment-based habitual dispositions of the will	Enacting personality in the real world	Implementing positive and negative values in all sort of goods and evils

## Bibliography

- Aristotle (1975): *On the Soul*. Cambridge: Harvard University Press.
- Bower, Matthew/Gallagher, Shaun (2013): "Bodily affects as prenoetic elements in enactive perception". In: *Phenomenology and Mind* 4, pp. 78–93.
- Calabi, Clotilde (1996): *Passioni e ragioni*. Milan: Guerini.

- Calabi, Clotilde/Di Francesco, Michele (2000): "Eventi (e cause) mentali. Il caso delle emozioni". In: Gabriele Usberti (ed.): *Modi dell'oggettività, Saggi in onore di Andrea Bonomi*. Milan: Bompiani.
- Damasio, Antonio R. (1994): *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putman & Sons.
- Damasio, Antonio R. (2010): *Self comes to mind. Constructing the conscious brain*, New York, Pantheon Books.
- Davidson, Donald (1980): *Action and Events*. Oxford: Oxford University Press.
- De Monticelli, Roberta (2000): *L'avenir de la phénoménologie – Méditations sur la connaissance personnelle*, Paris: Aubier-Flammarion.
- De Monticelli, Roberta (2013): "Requiredness. An argument for Value Realism". In: *Phenomenology and Mind* 5, pp. 84–97.
- De Monticelli Roberta (2003): *L'ordine del cuore – Etica e teoria del sentire*. Milano: Garzanti.
- De Sousa, Ronald (1987): *The Rationality of Emotions*. Cambridge: MIT Press.
- De Sousa, Ronald (1995): "The Body Is In". In: *The Semiotic Review of Books* 6 (2).
- Desmond, William (1995): *Perplexity and Ultimacy*. New York: SUNY Press.
- Fuchs, Thomas/Koch, Sabine C. (2014): "Embodied affectivity: on moving and being moved". In: *Frontiers of Psychology* 5, pp. 1–12.
- Gallagher, Shaun (2005): *How the Body Shapes the Mind*. Oxford: Clarendon Press.
- Gallagher, Shaun/Schmicking, Daniel (eds.) (2010): *Handbook of Phenomenology and Cognitive Science*. Dordrecht/New York: Springer.
- Gallagher, Shaun 2013, *Phenomenology*, Basingstoke: Palgrave Mcmillan.
- Gibson, James J. (1966): *The senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.
- Gibson, James J. (1986 [1979]), *The ecological approach to visual perception*, Boston: Houghton Mifflin.
- Goldie, Peter (2000): *The Emotions*. Oxford: Oxford University Press.
- Jones, Keith S. (2003): "What is an affordance?". In: *Ecological Psychology* 15(2), pp. 107–114.
- Magri, Tito (a cura di) (1999): *Filosofia ed emozioni*. Milan: Feltrinelli.
- Mc Dowell, John (1998): *Mind, Value and Reality*. Cambridge: Harvard University.
- Merleau-Ponty, Maurice (1945): *Phénoménologie de la perception*. Paris: Gallimard.
- Nussbaum, Martha (2001): *Upheavals of Thought: The Intelligence of Emotions*. Cambridge: Cambridge University Press.
- Oksenberg Rorty, Amy (1980): *Explaining Emotions*. Berkeley: Univ. of California Press.
- Scheler, Max (1980): *Formalism in Ethics and Non-Formal Ethics of Values*. Transl. by M. Frings and R.L. Funk. Evanston, IL: North Western University Press.
- Ratcliffe, Matthew (2008): *Feelings of Being – Phenomenology, Psychiatry and the Sense of Reality*. Oxford: Oxford University Press
- Rudder Baker, Lynne (2013): *Naturalism and the First-Person Perspective*. Oxford: Oxford University Press.
- Tappolet, Christine (2000): *Emotions et valeurs*. Paris: Presse universitaire de France..





Kathi Beier

# Stolz und Vorurteil. Über einige Schwierigkeiten der ethischen Selbstbewertung

**Abstract:** “Pride and Prejudice” addresses the question why it is that we human beings cannot know about our moral or ethical qualities in such a way that we are able to infallibly ascribe them to ourselves. Given that there is something like a first-person authority with respect to the knowledge of one’s own feelings, beliefs, and intentions, the lack of such an authority with respect to the knowledge of one’s virtues has to be explained. The difficulties in evaluating oneself ethically, e.g. to take oneself to be modest or temperate, as well as the claim that this can only be done by way of adopting another person’s perspective is perfectly presented in Plato’s *Charmides*. Plato, however, does not answer the question what the difficulty rests upon. After a careful exploration of the problem, the paper presents two possible answers. For both of them, Aristotle’s considerations in the *Nicomachean Ethics* play a crucial role.

## 1 Das Problem der ethischen Selbstbewertung

Mit Blick auf das Wissen um die eigene Person, so scheint es, gibt es eine klare Autorität der Erste-Person-Perspektive. Wer außer mir selbst sollte wissen und einschätzen können, was ich denke, fühle, beabsichtige etc.? Empfindungen sind dafür ein klassisches Beispiel. Sie können von anderen Personen zwar erkannt, aber nicht mitempfunden werden, so dass andere nicht wissen, was und wie ich empfinde. Zudem kann man seine Gefühle vor anderen verbergen. Diese Autorität der ersten Person Singular wird auch an anderen Phänomenen deutlich, etwa bei Überzeugungen. Ob Aussagen von anderen wahrhaftig oder gelogen sind, kann niemand außer der sprechenden Person selbst sicher wissen. Denn nur diese weiß, was sie für wahr hält und damit, ob das, was sie für wahr hält, mit dem übereinstimmt, was sie als wahr behauptet. Schließlich zeugt von einem privilegierten Zugang zum Wissen um die eigene Person auch die Praxis in kriminalistischen Kontexten. Ermittler und Richter drängen auf ein Geständnis des mutmaßlichen Täters bzw. der Täterin, weil sich meist erst dadurch die Tat vollständig aufklären lässt. Denn erst, indem man vom Täter erfährt, welche Absichten er verfolgte, kann man wirklich sicher bestimmen, welche Tat vorliegt. Über ihre Absichten kann aber in der Regel die handelnde Person selbst am besten Auskunft

geben. In zentralen Angelegenheiten des Wissens um die eigene Person scheint es also eine Autorität, ja ein Privileg der Ich-Perspektive bzw. der ersten Person Singular zu geben.

Andererseits fällt jedoch auf, dass dies mit Blick auf ein bestimmtes Wissen von sich offensichtlich nicht gilt. Dort nämlich, wo es darum geht, sich selbst ethisch einzuschätzen, z. B. Auskunft darüber zu geben, ob man eine bestimmte Tugend besitzt oder insgesamt tugendhaft ist, oder ob man etwas gut getan hat oder in etwas gut ist, fällt das Urteil ungemein schwer. Ja, vielleicht ist ein solches Urteil gerade einem selbst überhaupt nicht möglich. Das lässt sich an einigen Beispielen verdeutlichen. Klassisch ist der Fall einer Person, die betont, dass und wie bescheiden sie sei, und die gerade dadurch zeigt, dass sie nicht bescheiden ist. Dasselbe gilt mit Blick auf die Tugend der Demut. Allgemeiner gefasst: Wann immer jemand anderen gegenüber seine (angeblichen) Vorzüge herausstellt, ruft das bei den Hörern Missbehagen hervor. Man fragt sich: Kann diese Person das überhaupt beurteilen? Sollte sie eine solche Einschätzung nicht besser anderen überlassen? Dieser Eindruck verstärkt sich, wenn man z. B. Briefe und Aufzeichnungen von Heiligen liest, etwa die von Mutter Theresa. Dabei fällt nämlich auf, wie diese trotz der guten Taten, die sie vollbracht haben, sich immer wieder als Sünder empfinden, von ihren Fehlern sprechen und um die Kraft zur Besserung bitten. Das ist meiner Meinung nach ein Beleg dafür, dass offenbar nicht einmal solche Menschen, die von vielen als moralisches Vorbild empfunden werden, von ihrer eigenen ethischen oder moralischen Güte wissen. Wenn das aber so ist, dann ist die Frage, der ich hier nachgehen will, die Frage danach, warum das so ist. Was ist der Grund dafür, dass es *kein* Privileg der Erste-Person-Perspektive bei der ethischen Selbstbewertung gibt? Warum sind wir, wie es scheint, in dieser Hinsicht gewissermaßen „betriebsblind“ und deshalb eher auf andere angewiesen?

Bevor ich mögliche Antworten auf diese Frage diskutiere, möchte ich das zu klärende Phänomen und das Problem, das es meiner Ansicht nach aufwirft, zunächst noch einmal anders verdeutlichen, und zwar mit Blick auf Platons Dialog *Charmides*. Darin beginnt Sokrates mit Charmides – einem jungen Mann, von dem alle anderen Gesprächspartner übereinstimmend sagen, dass er nicht nur schön an Gestalt sei, sondern auch schön an der Seele, da er besonnen sei – ein Gespräch über Begriff und Bedeutung der Besonnenheit (*sophrosyne*). Indem Sokrates dieses Gespräch mit Charmides und nicht mit einem der anderen Anwesenden führt, unterstellt er, dass Charmides, insofern er besonnen sei, auch wissen müsse, dass er es sei, und dass er deswegen auch wissen müsse, was Besonnenheit sei. In diesem Sinne sagt Sokrates zu Charmides:

Offenbar nämlich, wenn dir die Besonnenheit (*sophrosyne*) beiwohnt, musst du auch etwas von ihr auszusagen wissen. Denn notwendig muss ihr Einwohnen (*enechein*), wenn sie dir

einwohnt, eine Empfindung (*aisthesis*) hervorbringen, auf welcher dir dann irgendeine Vorstellung (*doxa*) von der Besonnenheit sich gründet, was sie wohl ist und worin sie besteht. Oder meinst du nicht so? – Das meine ich wohl, sprach er. (Platon, *Charmides* 158e-159a)

Sokrates unterstellt hier also, so könnte man sagen, das eben angesprochene Privileg der ersten Person Singular. Demnach scheinen einem einerseits die eigenen Eigenschaften und Qualitäten doch in der Tat am nächsten zu sein oder wenigstens so nahe, dass man sie zuerst bemerken müsste, zumindest eher als andere. Das ist die erste von Sokrates unterstellte Verbindung, die von *enechein* und *aisthesis*. Wenn das aber so ist, dann bemerkt man offenbar auch, dass man diese Eigenschaften und Qualitäten hat, und hat dadurch zugleich einen privilegierten Zugang zu ihrer Bedeutung und begrifflichen Bestimmung. Das ist die zweite von Sokrates unterstellte Verbindung, die von *aisthesis* und *doxa* – auch wenn dieser zweite Punkt für meine folgenden Überlegungen nicht weiter relevant ist.

Andererseits bringt Sokrates Charmides mit seiner Frage, ob dieser besonnen sei, in eine schwierige Lage – und versteht diese Schwierigkeit dann auch sofort. Charmides weiß nämlich auf die Frage des Sokrates, ob er besonnen sei, keine rechte Antwort. Platon beschreibt es so:

Da errötete Charmides zunächst [...], dann antwortete er ohne jede Schüchternheit, indem er erklärte, es sei nicht leicht, so im Augenblick die Frage mit Ja oder mit Nein zu beantworten. Denn wenn ich, sagte er, leugne, besonnen zu sein, so würde es doch einerseits auffallend sein, so gegen sich selbst zu zeugen, andererseits aber würde ich unsern Kritias und dazu noch viele andere Lügen strafen, die mich für besonnen halten, wie sie behaupten. Erkläre ich mich aber für besonnen und lobe mich demnach, so wird das leicht anmaßlich erscheinen. Ich weiß also nicht, was ich dir antworten soll. – Darauf sagte ich [gemeint ist Sokrates]: Mir scheint das ganz billig, was du sagst, Charmides, und mich dünkt daher, wir sollten gemeinschaftlich untersuchen, ob du das besitzt oder nicht, wonach ich frage. (Platon, *Charmides* 158c-e)

Charmides erkennt die Zwickmühle, in der er sich befindet. Auf der einen Seite mutet es seltsam an, von sich nicht zu wissen, ob einem eine bestimmte ethische Qualität, in diesem Fall die Tugend der Besonnenheit, zukommt. Auf der anderen Seite ist es fast noch seltsamer, sich diese Qualität unverhohlen zuzuschreiben. Denn wie Charmides erkennt, bedeutet das, sich selbst zu loben. Und Selbstlob ist anmaßend. Das wiederum versteht Sokrates und bietet sofort seine Hilfe an. Er legt dadurch nahe, dass die eigene ethische Bewertung gerade *nicht* von einem selbst und alleine vorgenommen werden kann, und zwar offenbar deshalb, weil man hier nicht zu einem verlässlichen Wissen kommen kann. Aber warum eigentlich nicht?

## 2 Die Schueler-Driver-Kontroverse

Mit Blick auf die Tugend der Bescheidenheit gibt es eine interessante Debatte zwischen Julia Driver und G.F. Schueler. Driver vertritt die These, dass es so genannte „virtues of ignorance“ gebe,<sup>1</sup> dass also – entgegen der üblichen und gerade auch bei Platon vorfindlichen Annahme, dass Tugend mit Wissen zusammenhängt – manche Tugenden keine Tugenden wären, würden sie nicht mit Unwissenheit einher gehen. Das macht sie an der Tugend der Bescheidenheit deutlich. Hier gehe die Unwissenheit unmittelbar in das Definiens der Tugend ein, denn Bescheidenheit sei die Disposition, seinen Selbstwert in bestimmter Hinsicht und bis zu einem bestimmten Grad zu unterschätzen. Wer bescheiden ist, so Driver, sei unwissend in Bezug auf seinen Wert; er schätze sich selbst falsch ein, z. B. nicht als den größten lebenden Physiker, der man tatsächlich sei, sondern nur als den drittgrößten.

I have argued that the best account of modesty is an underestimation account. On this view, modesty involves – or at least can involve – an agent underestimating self-worth in some respect, to some limited degree. The degree of underestimation must be limited in order to differentiate modesty from a vice such as self-deprecation. [...] if Albert Einstein viewed himself as a great physicist, just not the greatest physicist of the twentieth century – that’s modesty. He is mistaken but not dramatically off the mark. (Driver 1999, p. 827)

Bescheidenheit beinhaltet Driver zufolge also eine Unwissenheit in Bezug auf sich selbst, die aber gut und daher eine Tugend sei, da sie die, so Driver wörtlich, „gegebene“ und „destruktive Tendenz“ von Menschen korrigiere, sich mit anderen zu vergleichen.<sup>2</sup> Diese Korrektur ist gut und schätzenswert mit Blick auf ihre Folgen für das soziale Miteinander, denn eine Person, die bescheiden ist, so Driver, wird andere Personen nicht zu neidischen Reaktionen provozieren.

The modest person is one who does not spend a lot of time ranking, who does not feel the need to do so, and thus remains ignorant to the full extent of self-worth (to a limited extent). The analogy with beauty is helpful [...] Someone who doesn’t compare his appearance to those around him, and, even better, seems unaware of it, seems less likely to provoke an envy response in others. (Driver 1999, p. 828)

Beides, die konsequentialistische Begründung des Werts der Bescheidenheit wie auch die Erläuterung ihrer Bedeutung, hat die – wie ich finde berechtigte – Kritik von G.F. Schueler hervorgerufen. Er verweist, was die Begriffsanalyse angeht, zum

---

<sup>1</sup> Vgl. Driver 1999 und 2001, pp. 16–41.

<sup>2</sup> Den Gedanken, dass Tugenden Korrektive sind, übernimmt Driver von Philippa Foot; vgl. Foot 1978.

einen darauf, dass es eine Spannung zwischen den beiden zentralen Gedanken von Driver gibt, nämlich zwischen dem Gedanken, dass die bescheidene Person jeden Vergleich und jedes Ranking vermeidet, und dem Gedanken, dass sie sich selbst in einem Ranking weniger hoch ansiedelt, als sie es verdiente. Man kann nicht beides zugleich behaupten. Zum anderen bestreitet Schueler rundheraus, dass derjenige, der sich irrtümlicherweise bloß als den drittgrößten Physiker, nicht als den größten betrachtet, überhaupt bescheiden sei. Bescheiden sei eine Person vielmehr dann, wenn sie indifferent sei gegenüber den Eindrücken, die andere von ihr und ihren Errungenschaften oder Leistungen haben. In den Worten von Schueler: „A modest person is a person who does not care at all whether people are impressed with her for her accomplishments.“ (Schueler 1999, p. 836) Gegen die These von Driver behauptet Schueler zudem, dass Bescheidenheit intrinsisch wertvoll sei. Dieser intrinsische Wert komme in den Blick, wenn man sich fragt, was an einer unbescheidenen Person eigentlich so abstoßend ist. Es ist, so seine These, die Tatsache, dass die Unbescheidenheit eine gewisse Hohlheit („hollowness“) der unbescheidenen Person offenbare, insofern diese nämlich ihre Ziele und Zwecke nicht aus sich heraus entwickle, sondern von den Leuten um sich herum, die sie beeindrucken will. Das heißt im Gegenzug: Bescheidenheit ist eine Tugend mit Blick auf das, was sie über die bescheidene Person offenbart, nämlich dass diese über eine Art Substanz verfügt und ihre Ziele und Zwecke aus sich selbst heraus gewinnt.<sup>3</sup>

Mir kommt es bei dieser Debatte zwischen Driver und Schueler mit Blick auf meine Fragestellung vor allem auf zwei Gedanken an. Der erste Gedanke betrifft den Zusammenhang von Bescheidenheit und Unwissenheit. Auch wenn Schuelers Analyse der Bescheidenheit anders als die von Driver keinen unmittelbaren Verweis auf Unwissenheit enthält, so kann man durch sie doch verstehen, warum Bescheidenheit rein begrifflich betrachtet mit Unwissenheit auf der Metastufe einher gehen muss, warum also eine bescheidene Person von sich nicht wissen kann, dass sie bescheiden ist. Denn wenn Bescheidenheit bedeutet, indifferent gegenüber dem Lob und den Urteilen anderer zu sein, dann zeigt derjenige, der von sich glaubt, bescheiden zu sein, und sich diese Tugend auch zuschreibt, gerade, dass er *nicht* bescheiden ist. Denn der Zweck dieser Selbstzuschreibung scheint zu sein, sich des Lobes und der Bewunderung anderer zu versichern – auch dann, wenn man die Zuschreibung nur still in Gedanken vornimmt. Man versichert sich damit, dass die anderen, auch wenn sie es faktisch nicht tun, einen schätzen und bewundern müssten.

---

<sup>3</sup> Für eine ähnliche Kritik an Driver, die zudem die tugendethischen Grundlagen stärker reflektiert, als Schueler dies tut, vgl. Winter 2012.

Damit hängt ein zweiter Gedanke zusammen. Denn das gerade Behauptete könnte nahe legen, dass ich meine generelle These von der Unmöglichkeit des unmittelbaren Wissens bzw. des Selbst-Wissens um die eigene Tugendhaftigkeit einschränken müsste, sie nur mit Blick auf bestimmte Tugenden aufrecht erhalten könne, nämlich nur mit Blick auf solche, deren unmittelbare Bedeutung eine Selbstzuschreibung ausschließen. Und das hieße, dass man eigentlich induktiv vorgehen müsste, d. h. von jeder einzelnen Tugend zunächst deren Bedeutung analysieren und dann schauen müsste, ob diese Bedeutung ein Selbstwissen und in der Folge eine Selbstzuschreibung begrifflich verhindert. Doch das scheint mir der falsche Weg zu sein, da offenbar von *jeder* Tugend gilt, dass ihr Selbstwissen und ihre Selbstzuschreibung anmaßend sind.<sup>4</sup> Sicher ist es bei der Tugend der Bescheidenheit so, dass die Person, die gefragt wird, ob sie bescheiden sei, und die darauf aufrichtig antworten will, in besondere Probleme gerät. Bei ihr scheint es eine Art performativen Selbstwiderspruchs zu geben, wenn sie rundheraus von sich behauptet, sie sei bescheiden.<sup>5</sup> Diesen internen Widerspruch gibt es bei anderen Tugenden vielleicht nicht. Und dennoch: Ist es nicht so, dass auch diejenigen, die von sich zu glauben wissen, dass sie klug, gerecht, tapfer etc. sind, und dies auch ohne Bedenken von sich behaupten, seltsam anmuten? Hier liegt vielleicht kein interner Widerspruch zwischen dem Gehalt der jeweiligen Tugend und der Selbstzuschreibung davon vor wie im Falle der Bescheidenheit. Aber es scheint dennoch so zu sein, dass eine Selbstzuschreibung nicht möglich ist. Eine tugendhafte Person sagt von sich überhaupt keine Tugend aus, egal, welche es ist. Warum nicht? Meine These ist, dass sie weiß, dass sie nicht darum wissen kann – oder nur sehr mittelbar und indirekt, nämlich über die Urteile anderer. Und dann kann sie das ethische Urteil über sich lieber gleich anderen überlassen.

Deswegen will ich deduktiv vorgehen und versuchen zu erklären, warum von allen Tugenden *qua* Tugenden gilt, dass derjenige, der sie hat, nicht unmittelbar um sie wissen kann – zumindest gegeben unsere gewöhnliche Natur als Menschen, hier v. a. unsere Urteilsnatur. Das, was alle Tugenden verbindet, ist ihr „dichter“ oder „dicker“ Charakter: Sie beschreiben nicht bloß die Haltungen einer Person, sondern sie bewerten sie zugleich.<sup>6</sup> Dieser evaluative Aspekt scheint der Grund für die Unmöglichkeit des unmittelbaren Wissens um sie zu sein.

---

4 Ich beschränke mich in diesem Text auf die Diskussion des Wissens um die eigene Tugendhaftigkeit und lasse die Frage, inwiefern das Wissen und die Selbstzuschreibung von Lastern ähnliche Probleme aufwerfen, offen.

5 Und zwar, wie Julia Driver richtig feststellt, einen Selbstwiderspruch „as in the case, for example, of someone saying ‚I don’t talk‘“ (Driver 2001, p. 17).

6 Vgl. Williams 1987, Kap. 7 und 8.

### 3 Selbstbewusstsein oder Selbsterkenntnis?

Die Gegenthese zu der eben formulierten vertritt Sebastian Rödl in seinem Buch *Selbstbewusstsein*. Rödl zufolge stimmt es schlicht nicht, dass man von seiner eigenen Tugendhaftigkeit nicht wissen kann. Im Gegenteil, man wisse notwendigerweise darum, nämlich insofern jeder von uns ein selbstbewusstes Wesen sei und sich dieses Selbstbewusstsein auch in unserem ethischen bzw. moralischen Handeln ausdrücke. Rödl schreibt:

Eine praktische Lebensform ist eine Einheit unendlicher Zwecke. Und die Vorstellung eines unendlichen Zwecks ist in dem Sinne praktisch, dass sie Handeln verursacht, das mit dem Zweck übereinstimmt. Jemandes Vorstellung der Gerechtigkeit etwa ist die Ursache seines gerechten Handelns, und das nicht zufällig, denn die Kausalität seiner Vorstellung ist ein Denken, das sein Handeln aus diesem Zweck ableitet. Sein Wissen, dass er so handelt, wie es die Gerechtigkeit erfordert, ist daher spontan; es schließt seinen Gegenstand ein und ist in ihm eingeschlossen.

(Rödl 2011, S. 239)

Dazu möchte ich zwei Bemerkungen machen. Rödl argumentiert hier, erstens, ganz klar kantisch. Dass absichtliches Handeln durch eine Kausalität des Denkens charakterisiert ist, soll denselben Gedanken ausdrücken, den Kant in der *Grundlegung zur Metaphysik der Sitten* so formuliert hat: „Da zur Ableitung der Handlungen von Gesetzen Vernunft erfordert wird, so ist der Wille nichts anderes als praktische Vernunft.“ (Kant, GMS 412) Dennoch, so scheint mir, ist sich Kant nicht so sicher wie Rödl, dass der gerecht Handelnde auch Wissen von seiner Gerechtigkeit hat. Vielmehr streitet Kant in der *Grundlegung* ein sicheres Wissen um die eigene Moralität gerade ab. Niemand könne selbst beurteilen, ob er wirklich „aus Pflicht“ oder nicht bloß „pflichtgemäß“ gehandelt habe, so Kant, da sich immer eine Neigung in die Handlungsmotivation eingeschlichen haben könne, die die moralische Quelle des Handelns verderbe. Kurz: Kant scheint mir nicht so eindeutig zu sein, wie Rödl ihn darstellt. Die zweite Bemerkung ist eher eine Frage: Ist es wirklich ein spontanes, selbstbewusstes Wissen, das man, insofern man um der Gerechtigkeit willen handeln will, dann auch von seiner eigenen Gerechtigkeit hat? Ich habe daran Zweifel. Rödl legt nahe, dass das Wissen um die eigene Tugendhaftigkeit insofern selbstbewusstes Wissen sei, als tugendhafte Handlungen absichtliche Handlungen sind; und das könnte sie nur sein, wenn sie das Resultat einer praktischen Überlegung sind. (Soweit kann auch jeder Platoniker und jeder Aristoteliker mitgehen.) Wenn nun, so die Unterstellung Rödl's, meine Tugendhaftigkeit Teil meiner praktischen Überlegung ist, dann ist auch das Wissen um meine Tugendhaftigkeit spontanes, selbstbewusstes Wissen.



Doch dieser Schluss scheint mir übereilt, denn es sprechen aus meiner Sicht zwei Gründe gegen ihn.

Der erste Grund hat mit der möglichen Diskrepanz zwischen Absicht, Akt und Tugendhaftigkeit zu tun. Die praktische Überlegung, die dem absichtlichen Handeln vorausgeht, stellt den Beweggrund meines Handelns dar, sie macht mein Motiv explizit. Aber sie garantiert nicht, dass das, was ich dann wirklich tue, auch dem entspricht, was ich wollte und beabsichtigt habe. Die beste Absicht kann in einem schlechten Handeln resultieren. Weil man nicht immer alle Umstände der Handlung kennt oder viele davon unserem Einfluss entzogen sind, kann die beste Absicht beim Versuch ihrer Ausführung in einer schlechten Tat enden.<sup>7</sup> Die Beurteilung der Tugendhaftigkeit einer Person hängt nun aber nicht von der Absicht des Handelnden alleine ab, sondern auch von ihren Akten. Und da das Gelingen der Umsetzung meiner guten Absicht in gute Akte durch die Absicht allein nicht garantiert werden kann, kann ich von meiner Tugendhaftigkeit auch kein selbstbewusstes Wissen haben. Und selbst da, wo Absicht und Akt zusammen fallen, wo man also genau das tut, was man wollte, kann es eine Diskrepanz zwischen guter Absicht und gutem Akt auf der einen Seite und der Tugendhaftigkeit der handelnden Person auf der anderen Seite geben. Daran erinnert Aristoteles im siebten Buch der *Nikomachischen Ethik*, wenn er die *akrasia*, also die Willensschwäche diskutiert. Dabei macht er nämlich auf den Unterschied zwischen der bloß beherrschten Person (dem *enkrates*) und der tugendhaften, d. h. mäßigen Person (*sophron*) aufmerksam. Beide, so Aristoteles, wollen das Gute und handeln auch gut, d. h. ihre Absichten und Akte sind vollkommen identisch. Aber die bloß beherrschte Person ist noch nicht tugendhaft, weil sie das Gute noch gegen innere Widerstände und Bedenken ausführt. Sie tut es – das unterscheidet sie vom *akrates* –, aber sie tut es nicht gern. Ihre Unlust ist für Aristoteles ein Zeichen dafür, dass sie die relevante Tugend, die Tugend der Mäßigkeit, noch nicht besitzt.<sup>8</sup> Das Wissen um meine Tugendhaftigkeit kann also kein selbstbewusstes, spontanes Wissen sein, weil weder aus meiner praktischen Überlegung bzw. meiner Handlungsabsicht alleine noch aus meinem tatsächlich guten Handeln folgt, dass ich tugendhaft bin.

Der zweite Grund macht auf die strukturelle Schwierigkeit einer vorurteilslosen Perspektive aufmerksam. Mir scheint nämlich eine weitere Überlegung von Aristoteles wichtig zu sein, aus der folgt, dass und warum jede unmittelbare oder

---

<sup>7</sup> Viele anschauliche Beispiele dafür finden sich im dritten Buch von Aristoteles' *Nikomachischer Ethik*, vgl. insbesondere 1111a7 – 15.

<sup>8</sup> Kant scheint dagegen zu sagen, wer das Richtige ungern tut, ist moralisch besser, weil moralisch eindeutiger, denn er handelt wirklich nur aus Pflicht, nicht aus Neigung. Aber das ist und war seit jeher ein schwer verständlicher Gedanke.

direkte Selbstbeurteilung in ethischen Fragen so schwierig ist.<sup>9</sup> Aristoteles diskutiert im sechsten Buch der *Nikomachischen Ethik* unter anderem das Verhältnis von Klugheit (*phronesis*) und Tugendhaftigkeit bzw. Gutheit des Charakters (1144a30-b32). Beide sind ihm zufolge aufs Engste miteinander verbunden. Denn einerseits kann nur der Kluge tugendhaft sein, weil Tugendhaftigkeit die Disposition ist, verbunden mit einer richtigen Überlegung zu handeln; die richtige Überlegung über solche Dinge aber ist die Klugheit. Das begründet die oben erwähnte Ansicht, dass Tugend und Unwissenheit nicht gut zusammen passen. Andererseits – und dieser Gedanke ist für die Frage der ethischen Selbstbeurteilung entscheidender – ist Klugheit an Tugendhaftigkeit gebunden, weil eine praktische Überlegung nur dann richtig sein kann, wenn ihr Ausgangspunkt, also der Obersatz richtig ist. Dieser Ausgangspunkt (*archē*) zeige sich aber, so Aristoteles, nur dem tugendhaften bzw. guten Menschen. Deshalb sei es „offenkundig unmöglich, klug zu sein, wenn man nicht gut ist“. Der Gedanke, der dieser Argumentation zugrunde liegt, ist der Gedanke, dass unser Charakter unser praktisches bzw. ethisches Urteilsvermögen beeinflusst. Deshalb kann nur der ethisch gut urteilen, der einen guten Charakter hat. Wenn das aber so ist, wenn also der Charakter grundsätzlich das ethische Urteilen beeinflusst, dann gleicht der Versuch einer Person, ihren Charakter ethisch zu beurteilen, dem Versuch einer Katze, ihren Schwanz einzufangen: Man findet einfach keinen neutralen Standpunkt, von dem aus sich die Beurteilung angemessen betreiben ließe, denn man hat immer ein positives Vorurteil sich selbst gegenüber, immer einen *bias*.<sup>10</sup> Selbst tugendhafte Menschen wie Charmides, so führt Platon vor, scheitern an der Aufgabe, sich selbst ethisch richtig zu beurteilen. Sie scheitern, wie sich den Überlegungen von Aristoteles entnehmen lässt, aus strukturellen Gründen. Denn uns fehlt die Distanz zu uns selbst, die nötig ist, um zu einem unvoreingenommenen Urteil über uns und die Güte unserer Handlungen und Eigenschaften zu kommen. Die Distanz, die nötig ist, ist dabei vor allem eine affektive: Es geht im Sinne von Adam Smith vor allem darum, einen unparteiischen und interesselosen Blick auf sich selbst zu gewinnen. Das scheint aus den genannten Gründen anderen eher möglich zu sein als uns selbst.

---

<sup>9</sup> Ich beziehe mich hier auf meine Überlegungen in Beier 2010, S. 140.

<sup>10</sup> So sagt Aristoteles im neunten Buch der *Nikomachischen Ethik*, dass jeder Mensch und gerade jeder gute Mensch sich selbst lieben muss (1169a12); dabei grenzt er diese gute Art der Selbstliebe von dem ab, was er Selbstliebe im schlechten Sinne nennt, d. h. von einem reinen Egoismus.

## 4 Aporie: Stolz und Vorurteil

Wenn das allerdings so ist, dann muss man, wie es scheint, mit Aristoteles gegen Aristoteles argumentieren. Denn dann entsteht ein Problem mit derjenigen Tugend, die Aristoteles *megalopsychia* nennt, ins Deutsche übersetzt als groß an der Seele sein, Hochgesinntheit oder Stolz. Der Stolze zeichnet sich nämlich gerade dadurch aus, dass er sich in seinem Wert und seiner Würde richtig einschätzen kann. So heißt es bei Aristoteles:

Als stolz (*megalopsychos*) gilt, wer sich selbst großer Dinge für wert hält und dies auch [wirklich] ist. Wer sich ihrer für wert hält, ohne es zu sein, ist dumm; von denen aber, die Träger einer Tugend sind, ist keiner dumm oder ohne Verstand. [...] Der Stolze aber wird, wenn er tatsächlich der größten Dinge wert ist, der Beste (*aristos*) sein; denn der Bessere ist immer größerer Dinge wert und der Beste der größten. Der wahrhaft Stolze muss also gut (*agathos*) sein. Und Kennzeichen des Stolzen dürfte die Größe in jeder Tugend sein. [...] Es scheint also der Stolz eine Art Schmuck (*kosmos*) der Tugenden zu sein; denn er macht sie größer und besteht nicht ohne sie. Deswegen ist es schwierig, wahrhaft stolz zu sein; es ist nicht möglich, ohne ein edler und guter Mensch zu sein (*kalogagathia*). (Aristoteles, NE IV 7, 1123b2ff.)

Die Selbsterkenntnis, die der Stolze mit Blick auf seine ethische Güte hat, wird besonders deutlich, wenn Aristoteles die beiden entsprechenden Laster dieser ethischen Tugend beschreibt:

Wer den Mangel aufweist, ist kleinmütig (*mikropsychos*), wer das Übermaß, eitel (*chaunos*). [...] Denn der Kleinmütige, der guter Dinge wert ist, beraubt sich gerade der Dinge, deren er wert ist, und erweckt den Eindruck, darin einen Fehler zu haben, dass er sich selbst dieser guten Dinge nicht für wert hält und sich selbst nicht kennt; denn sonst würde er die Dinge erstreben, derer er wert ist, vorausgesetzt, dass sie gut sind. [...] Die eitlen Menschen andererseits sind dumm und kennen sich selbst nicht, und dies ganz offensichtlich; denn sie versuchen sich an ehrenvollen Dingen, wie wenn sie ihrer wert wären, und werden dann ihrer Unfähigkeit überführt. (Aristoteles, NE IV 9, 1125a17ff.)

Die Frage, die bleibt, ist also: Wenn die angeführten Gründe überzeugen, die dafür sprechen, dass wir selbst tendenziell nicht in der Lage sind, ein sicheres Wissen um unsere ethische Güte zu erlangen, wie ist dann die Tugend des Stolzes möglich? Was kann der Stolze, was die anderen nicht können? Wie schafft er es, die erwähnten Schwierigkeiten zu überwinden? Kurz: Wie können und sollen wir die Disposition der *megalopsychia* verstehen? Aristoteles spricht, wenn er den Stolzen charakterisiert, viel von Unabhängigkeit bzw. von einem Unbeeindruckt-sein von der Ehrerbietung durch andere, da der Stolze weiß, dass er größer, ja der größten Dinge wert ist, es aber „keine Ehre [gibt], die der vollkommenen Tugend angemessen ist“ (1124a7). Das lässt mindesten zwei Verständnismöglichkeiten zu.

Die Betonung könnte einerseits auf der Unabhängigkeit von der Ehrung durch andere liegen. Diese Unabhängigkeit ist gut und tugendhaft, weil sie eine Eigenständigkeit der Person demonstriert – ähnlich wie G.F. Schueler dies mit Blick auf die bescheidene Person behauptet. Man lässt sich in seinen Wünschen, Handlungen und Entscheidungen nicht von der oft schwankenden Meinung anderer beeinflussen, verlangt nicht nach deren Lob und Bewunderung. Diesem Verständnis zufolge geht es bei der Tugend des Stolzes vor allem darum, richtig mit der Ehre als dem größten äußeren Gut umzugehen. Das ist die unproblematischere Deutung für die in diesem Text entwickelte These, da sie nicht das Wissen betont.

Die Betonung könnte andererseits gerade auf dem Wissen um den eigenen Wert liegen, auf dem Wissen um die eigene Tugendhaftigkeit, deren „Krone bzw. Schmuck“ der Stolz ist. Wenn aber dieses Wissen die Voraussetzung für die Tugend des Stolzes ist, dann ist Stolz nach meiner Argumentation für uns Menschen höchst ungewöhnlich, da uns dieses Wissen, gegeben unsere Natur, nicht unmittelbar möglich ist.

## Bibliographie

- Aristoteles (2006): *Nikomachische Ethik*. Übersetzt und herausgegeben von Ursula Wolf. Reinbek bei Hamburg: Rowohlt.
- Beier, Kathi (2010): *Selbsttäuschung*. Berlin/New York: Walter de Gruyter.
- Driver, Julia (1999): „Modesty and Ignorance“. In: *Ethics* 109, pp. 827–834.
- Driver, Julia (2001): *Uneasy Virtue*. Cambridge: Cambridge University Press.
- Foot, Philippa (1978): *Virtues and Vices*. Oxford: Basil Blackwell.
- Platon (2011): *Charmides*. In: Werke in acht Bänden, 1. Band, herausgegeben von Gunther Eigler, übersetzt von Friedrich Schleiermacher. Darmstadt: Wissenschaftliche Buchgesellschaft, S. 287–349.
- Schueler, G.F. (1999): „Why is Modesty a Virtue?“ In: *Ethics* 109, pp. 835–841.
- Rödl, Sebastian (2011): *Selbstbewusstsein*. Berlin: Suhrkamp.
- Williams, Bernard (1985): *Ethics and the Limits of Philosophy*. Cambridge, Mass.: Cambridge University Press.
- Winter, Michael Jeffrey (2012): „Does Moral Virtue Require Knowledge? A Response to Julia Driver“. In: *Ethical Theory and Moral Practice* 15, pp. 533–546.



# Index of Names

- Abela, Paul 146  
Adorno, Theodor W. 145, 80f.  
Alexander, Bruce K. 210  
Alexander, Joshua 4  
Alharoun, Mohammad 363, 366  
Allais, Lucy 146  
Allison, Henry 146  
Althusser, Louis 83  
Anscombe, Elizabeth 2, 39, 53, 65, 255, 310  
Apel, Karl Otto 81  
Aquinas, Thomas 247–257, 293, 304  
Aristotle 56f., 59–61, 81, 135, 157f., 161, 195, 204, 206, 215, 289–295, 297–300, 302–306, 384, 389, 401, 408–410  
Audi, Robert 371f.  
Austin, John Langshaw 82–84, 310  
Avramides, Anita 311, 319  
Ayer, A. J. 1
- Bacharach, Michael 286  
Bachelard, Gaston 203, 206, 210–213  
Bacon, Francis 113, 116, 123  
Bakan, Joel 261  
Baker, Lynne R. 263, 396  
Banham, Gary 146  
Barbaras, Renaud 242  
Barnes, Barry 34f.  
Bealer, George 129f.  
Ben-Ze'ev, Aaron 266, 271  
Bennett, Jonathan 144  
Bennett, Maxwell 85f., 199  
Benson, Paul 323, 329–331  
Bergson, Henri 79, 212, 221  
Bernet, Rudolf 147  
Bird, Graham 146  
Blackmore, Susan 234  
Bloor, David 34f.  
Blumberg, Philip I. 263  
Boddy, Clive R. 269  
Boghossian, Paul 31f., 34, 41  
Bongard, Josh 218  
Boring, Edwin G. 122  
Bouwsma, Oets Kolk 66
- Bratman, Michael E. 83, 302f.  
Bratton, William W. 263  
Brentano, Franz 96f., 113, 129, 131–133, 135–140, 144f., 389  
Brey, Philip 124  
Briggs, Rachael 263, 271  
Brinkmann, Svend 199  
Bruder, Martin 270  
Byrd, Sharon B. 356
- Canfield, John 317  
Carnap, Rudolf 2, 79  
Carr, David 323  
Cassirer, Ernst 81, 114, 120  
Castoriadis, Cornelius 345  
Cavell, Stanley 81, 309f., 312f., 316, 320  
Chapman, James 220f.  
Chelstrom, Eric 323  
Chong, Jane 125  
Christman, John P. 323f.  
Chudnoff, Elijah 130, 140, 157  
Church, George 119  
Churchland, Patricia 83  
Churchland, Paul 83  
Clark, Andy 83, 126, 149f., 152  
Clifford, William K. 232, 234  
Cobb-Stevens, Robert 131  
Cohen, Michael 232  
Coliva, Annalisa 33, 41  
Collins, Randall 114, 122  
Colombetti, Giovanna 270  
Cooper, John M. 290, 300  
Coplan, Amy 194  
Copp, David 271  
Crane, Tim 83  
Crisp, Roger 291f., 295–297, 299, 301, 306  
Crowell, Steven 109, 330  
Cudworth, Ralph 113  
Cummins, Robert 130  
Curley, Edwin 352
- Damasio, Antonio 270, 381, 389, 396  
Dan-Cohen, Meir 263

- Darwin, Charles 119, 123  
 Davidson, Donald 2, 121, 255, 384, 389  
 Davis, Creston 80  
 De Biran, Maine 221  
 de Haan, Sanneke 221  
 De Jaegher, Hanne 223  
 De Vos, Jan 199  
 Deleuze, Gilles 2, 83, 118  
 Dennett, Daniel 81, 95, 97–105, 107, 232, 234  
 Deonna, Julien A. 271  
 DePaul, Michael 130  
 Depraz, Natalie 189  
 Derrida, Jacques 2, 4, 81, 118  
 Descartes, René 81, 113, 116, 137, 157–159, 165–170, 234, 315, 343, 381, 389, 396  
 Descombes, Vincent 117  
 Dewey, John 123  
 Dilthey, Wilhelm 121  
 Dressler, Joshua 356 f.  
 Dreyfus, Hubert 215  
 Driver, Julia 404–406  
 Drummond, John 324 f.  
 Drury, Maurice O'Connor 64  
 Duff, R. A. 352, 354  
 Dummett, Michael 1, 4, 79, 82, 112, 145
- Einstein, Albert 114, 404  
 Elster, Jon 112  
 Emerson, Ralph Waldo 124
- Fechner, Karl 123  
 Feigl, Herbert 233 f.  
 Ferrier, James 115  
 Fichte, Johann Gottlieb 114 f.  
 Finkelstein, David 314  
 Finnis, John 263  
 Fletcher, George 352  
 Føllesdal, Dagfinn 145  
 Foot, Philippa 404  
 Foucault, Michel 2, 118  
 Frankfurt, Harry 252, 323  
 Freer, Cameron E. 120  
 Frege, Gottlob 2, 4, 6, 11–15, 18 f., 22, 27 f., 65, 79, 82, 114, 120, 145, 396  
 French, Peter A. 263 f.  
 Freud, Sigmund 87, 389
- Frith, Uta 224  
 Fukuyama, Francis 117
- Gadamer, Hans-Georg 2, 4, 86, 121  
 Gallagher, Shaun 148, 153, 188, 195, 198, 223 f.  
 Gehlen, Arnold 124  
 Gendler, Tamar Szabó 204 f.  
 Georgen, Chris 100, 109  
 Gibbons, Sarah L. 146  
 Gibson, James 216, 386  
 Giddens, Anthony 121  
 Gilbert, Felix 338, 342  
 Gilbert, Margaret 262, 265, 302  
 Gilbert, Paul 83  
 Glock, Hans-Johann 81  
 Goethe, Johann Wolfgang von 65  
 Goetz, Bernhard 351, 358, 360–362, 366 f.  
 Goldie, Peter 193–195, 266 f., 271, 381–387, 389, 395 f.  
 Goris, Harm 250 f., 256  
 Graybiel, Anne 203, 205  
 Grayling, Anthony 31  
 Greenawalt, Kent 358  
 Griffin, David 234  
 Gross, Neil 112
- Hacker, Peter M. S. 47, 52, 61, 81–83, 85–87, 199  
 Hacking, Ian 112  
 Hager, Mark M. 263  
 Haller, Rudolf 31  
 Hanna, Robert 146  
 Hanson, Norwood Russell 3  
 Happé, Francesca 224  
 Harbisson, Neil 125  
 Hare, Richard M. 79 f.  
 Harman, Gilbert 34  
 Harnack, Adolf von 116  
 Hatfield, Gary 123  
 Heavey, Christopher L. 189  
 Hegel, Georg Wilhelm Friedrich 114 f., 117  
 Heidegger, Martin 2 f., 5, 79, 83, 86, 113, 144, 194 f., 216 f., 369  
 Heinrichs, Bert 372, 374  
 Helm, Bennett 193, 267  
 Henry, Michel 89, 240

- Hess, Kendy M. 263, 265, 271  
 Hildebrand, Dietrich von 369  
 Hindriks, Frank 263  
 Hirst, Paul Q. 263  
 Hitz, Zena 303 f.  
 Hobson, Peter 222, 224  
 Honneth, Axel 271  
 Huebner, Bryce 262 f., 272  
 Humboldt, Wilhelm von 114  
 Hurlburt, Russell 187, 189 f., 197  
 Husserl, Edmund 2 f., 6, 79, 81, 83, 85–87,  
 96 f., 107, 114, 120, 129, 131–135,  
 138–140, 144–147, 150–152, 157–159,  
 168–171, 188, 221, 233, 238, 240, 242,  
 272, 313–315, 317, 324–329, 369, 385  
 Hutto, Daniel 83  
  
 Illouz, Eva 262, 269  
 Ireland, Paddy 263  
  
 Jack, Anthony 189  
 Jacqueline, Dale 318  
 James, William 47, 56, 63, 112, 211, 219,  
 232  
 Johannsen, Wilhelm 119  
 Jung, Matthias 220  
 Jureschi, Livia Andreia 317  
  
 Kahn, Charles K. 298  
 Kahneman, Daniel 268  
 Kant, Immanuel 48 f., 52–54, 56–61, 81,  
 85, 114–116, 118 f., 144–147, 157–166,  
 204, 237 f., 369–377, 389, 407 f.  
 Kantorowicz, Ernst H. 261  
 Kapp, Ernst 124  
 Killminster, Suzy 324  
 Kind, Amy 150  
 Klimchuk, Dennis 355  
 Koffka, Kurt 87  
 Köhler, Wolfgang 63  
 Kojève, Alexandre 117  
 Körner, Stephan 144  
 Korsgaard, Christine M. 83, 262, 264, 330  
 Kosman, Aryeh 300  
 Koyré, Alexandre 117  
 Kraft, Julius 144  
 Kramer, Roderick M. 269  
  
 Kraus, Oskar 137  
 Kriegel, Uriah 83, 87–90  
 Krueger, Joel 195  
 Kuhn, Thomas 3, 119  
 Kung, Peter 8  
 Kusch, Martin 115, 263  
  
 Laclau, Ernesto 342, 345  
 Langton, Rae 146  
 Latour, Bruno 118  
 Laudan, Larry 119  
 Leder, Drew 221, 226  
 Leist, Anton 253  
 Leiter, Brian 81  
 Levinas, Emmanuel 369, 376 f.  
 Lipinska, Veronika 120  
 List, Christian 263 f., 278  
 Locke, John 58, 85, 113, 116, 122  
 Longuenesse, Beatrice 146  
 Luft, Sebastian 147  
 Lyotard, Jean-François 118  
  
 McCormick, Neil 358  
 Machiavelli, Niccolò 337–347  
 Mahoney, Martha R. 364  
 Makkreel, Rudolf A. 146  
 Malcolm, Norman 65 f., 70  
 Manning, Rita 263, 265  
 Mark, Gregory A. 263  
 Marty, Anton 137  
 Marx, Karl 122, 345  
 Massey, Gerald 4  
 Mathiesen, Kay 264  
 McDowell, John 81, 188  
 McIntyre, Ronald 145  
 McLuhan, Marshall 124  
 Mele, Alfred R. 267  
 Meltzoff, Andrew 222  
 Menary, Richard 83  
 Mendlow, Gabe 363, 366 f.  
 Mercolli, Laura 310  
 Merleau-Ponty, Maurice 1, 83 f., 87, 145,  
 188, 196, 203, 206–213, 217, 219, 231,  
 240–243, 253, 385  
 Merz, John Theodore 115  
 Meyers, Diana Tietjens 323  
 Minkowski, Eugène 220



- Moore, G.E. 37–39, 41, 47, 49–54, 65, 70,  
146, 222, 369, 371
- Moran, Dermot 147
- Moran, Richard 187, 191–195, 314 f., 318
- Mouffe, Chantal 342, 345
- Naberhaus, Thane M. 147
- Nagel, Thomas 81, 233, 247 f., 250, 254 f.,  
361 f.
- Nancy, Jean-Luc 346
- Natorp, Paul 145
- Neuhäuser, Christian 265
- Newton, Isaac 60, 113
- Noë, Alva 83, 196,
- Nordhaus, Ted 120
- Nussbaum, Martha C. 270
- Oshana, Marina A. L. 323
- Ostwald, Wilhelm 120
- Overgaard, Søren 80, 84, 313, 315
- Ozar, David T. 263 f.
- Parfit, Derek 262
- Pascal, Blaise 113
- Pavlov, Ivan 87
- Peacocke, Christopher 81
- Peirce, Charles Sanders 23, 119, 232
- Pettimengin, Claire 189
- Pettit, Philip 263 f., 277 f.
- Pfeifer, Rolf 218
- Piaget, Jean 87
- Platon 157 f., 388, 394, 401–404, 409
- Plessner, Helmuth 319
- Pocock, John G. A. 338
- Polanyi, Michael 216, 219, 226
- Popper, Karl 113, 119
- Porter, Theodore 123
- Prado, Carlos G. 1
- Price, Anthony W. 292, 294 f., 300
- Prichard, Harold A. 369 f.
- Prinz, Jesse J. 269 f.
- Pritchard, Duncan 33, 41
- Proctor, Robert 115
- Pust, Joel 130
- Quine, Willard Van Orman 2, 84, 121, 123
- Quinton, Anthony 1, 80
- Ramsey, William 130
- Rancière, Jacques 345
- Ranken, Nani L. 265
- Ravaisson, Felix 203, 206–211
- Rawls, John 369, 371, 373
- Regis, Ed 119
- Reiner, Hans 369
- Ricœur, Paul 203, 206, 210–214, 255
- Rinofner-Kreidl, Sonja 109, 324 f.
- Rödl, Sebastian 407
- Roepstorff, Andreas 189
- Roling, Bernd 256
- Rorty, Amelie Oksenberg 266
- Rorty, Richard 2, 112, 122
- Rosen, Gideon 34, 36
- Ross, William David 57, 369–372, 374
- Rovane, Carol, A. 262, 264 f., 278, 294
- Rowlands, Mark 83, 148
- Runciman, David 263
- Russell, Bertrand 2, 4, 79, 112, 116, 122 f.,  
146, 215, 232, 234, 240
- Ryle, Gilbert 82, 84, 204–208, 212 f., 215,  
218, 312
- Salmela, Mikko 265
- Santayana, George 123
- Sartre, Jean-Paul 2, 89, 117, 150 f., 193,  
195, 226, 272, 369
- Schacter, Daniel L. 221
- Scheler, Max 83, 157 f., 266, 277–285,  
287, 313, 369, 393 f.
- Schelling, Friedrich Wilhelm Joseph 114 f.
- Schleiermacher, Friedrich 121
- Schmid, Hans Bernhard 262, 265, 268 f.,  
284, 302, 306
- Schmitt, Frederick. F. 34, 367
- Schnädelbach, Herbert 115, 121
- Schönecker, Dieter 373 f.
- Schrödinger, Ernst 119, 237
- Schueler, George F. 404 f., 411
- Schulting, Denis 146
- Schütz, Alfred 83
- Schweikard, David P. 302
- Schwitzgebel, Eric 95, 101–104, 107–109,  
130, 189
- Scruton, Roger 264

- Searle, John 11–15, 20 f., 23, 26, 28, 81,  
83, 282, 285, 302
- Shakespeare, William 339
- Shapiro, Lawrence 149
- Shear, Jonathan 189
- Sheehy, Paul 263
- Shellenberger, Michael 120
- Sherman, Nancy 299 f.
- Sidgwick, Henry 369
- Smart, J. J. C. 51, 80 f.
- Smith, Adam 409
- Smith, Angela M. 331
- Smith, David W. 145
- Smith, Kyle 203, 205
- Smith-Pangle, Lorraine 291, 295
- Sokolowski, Robert 131
- Sokrates 168, 402 f.
- Solomon, Robert C. 114, 266, 270
- Sosa, Ernest 44
- Spengler, Oswald 124
- Spinoza, Baruch 53, 117, 204, 345
- Stanley, Jason 215
- Stein, Edith 83, 313
- Steinbock, Anthony J. 266, 323
- Stern, Daniel 223, 373 f.
- Stern, Robert 223, 374
- Stoljar, Natalie 323
- Stoll, Mary L. 263
- Strawson, Galen 87–90, 232–234, 236 f.
- Strawson, Peter F. 11, 13, 15–17, 19, 21–25,  
27 f., 44 f., 61, 88–90, 144, 254 f., 263,  
331, 369
- Stroud, Barry 41
- Stubenberg, Leopold 100
- Summa, Michela 257, 310, 314 f.
- Szanto, Thomas 262, 264 f., 268, 270,  
272, 287
- Taylor, Charles 84, 198
- Ter Hark, Michel 312–314, 317
- Teroni, Fabrice 271
- Thompson, Evan 196, 216
- Thomson, Judith Jarvis 34
- Titchener, Edward 102
- Torribio, Josefa 149 f.
- Tuomela, Raimo 83, 265, 302 f.
- Tye, Michael 81
- van Breda, Herman L. 84
- Varela, Francisco 189, 196, 216, 239
- Vargas, Manuel 267
- Velleman, David J. 303
- Verburgt, Jacco 146
- Wahl, Jean 84
- Waldenfels, Bernhard 315 f.
- Wallon, Henri 87
- Walther, Gerda 83
- Watson, John B. 87
- Wellman, Christopher C. 265
- Welton, Donn 147
- Wendt, Alexander 263
- Westlund, Andrea C. 329
- Whewell, William 119
- Williams, Bernard 41 f., 81, 84, 406
- Williams, Glanville 352
- Williams, Michael 32–36, 39–42, 44
- Wilson, Margaret 149
- Wiltsche, Harald 131
- Winch, Peter 257
- Winter, Michael Jeffrey 405
- Wissner-Gross, Alexander D. 120
- Witherspoon, Edward 311
- Wittes, Benjamin 125
- Wittgenstein, Ludwig 2, 4, 11–34, 37–44,  
47–60, 63–76, 81, 109, 187 f., 191, 257,  
309–314, 316–320
- Wolfe, Tom 117
- Wood, Allen 373
- Wright, Crispin 30, 32, 40
- Wright, Georg Henrik von 50
- Zabala, Santiago 80
- Zahavi, Dan 87, 89 f., 147, 188, 223, 268,  
279, 313



# Index of Subjects

- action 34–36, 47, 49–60, 83, 98, 125, 149, 192, 204f., 207–221, 223, 225–227, 236, 247, 252f., 255–257, 263f., 268, 272, 278, 280f., 287, 289f., 292–296, 300–305, 309, 311, 318, 325, 328f., 339, 345, 354, 364f., 382f., 385, 387, 389f.
- theory of 83, 247f., 254f.
- activity 20, 56, 137f., 148–152, 191f., 196, 207, 216, 237, 239, 243, 251f., 289–291, 293–296, 298–303, 305f., 310, 312, 346
- agency 111, 113, 187, 191–199, 204, 253, 256, 262–265, 267, 269, 289–292, 294–296, 298, 301f., 304, 306, 329, 331
- akrasia 408
- analysis 4, 82, 86, 97, 107, 130, 140, 213, 220f., 242
- analytic/continental divide 1–6, 79–91, 95, 111–114, 143f., 147–150, 231, 233
- analytic phenomenology 95f., 99, 101–104, 107
- analytic philosophy 1–6, 79–91, 140, 146, 231, 233, 381
- angel 247–257
- a priori, apriorism 4, 85, 97, 114, 118f., 160–164, 237f., 311, 345
- argument 4, 24, 52f., 80, 89, 95–97, 144, 157f., 169f.
- attention 108, 133, 136–139, 207, 209f., 218, 226f., 238, 241
- authority 96–98, 100, 104, 115, 117, 181, 191, 283f., 401
- autism 215, 223f.
- autonomy 199, 296, 323f., 327, 330
  
- behavior 34f., 148, 204, 210f., 218, 221, 223–235, 257, 261, 267, 270, 280f., 309, 311f., 314, 318, 381, 385f., 387, 390f., 398
- behaviorism 122, 148, 205, 243
- belief 29–36, 39, 42, 44, 97–105, 129f., 132, 140, 174–179, 181, 183, 191f., 278, 280, 283, 323, 329–332, 352f., 357, 359, 361–367, 384, 389, 401
- body 12, 16, 18, 29, 38, 148, 151, 208–210, 212–214, 215–223, 225–227, 240–242, 249, 252–255, 314, 382, 386, 396, 398
  
- Cartesianism 114, 145f., 187, 190, 210, 234, 315
- causality 47–49, 52f., 57f., 60
- certainty 29–45, 58, 65f., 70–74, 118, 234, 310–312, 314–317
- Charmides (Plato) 401–403, 409
- cognitive science 143, 147f., 188, 196
- collective intentionality 83, 262, 277, 281, 285f., 302f.
- color 22–27, 135–137, 65f., 71–73, 240
- common sense 86, 123, 289, 304f.
- communication 217, 220, 223, 225, 255–257, 315
- consciousness 6, 12–15, 20, 84, 87, 89f., 95–98, 100, 102, 104, 106–108, 113, 133, 138–140, 144, 147–149, 193f., 196, 206, 208f., 214, 217, 219, 231f., 234–236, 238, 240, 247f., 251, 255, 272, 279, 282, 304–306, 315, 317, 331, 382–385, 398
- continental philosophy 1–6, 79–83, 91, 95, 111–114, 143f., 148, 233
- contingency 36, 42–44, 337f.
- correlation 85, 269, 272, 311
- correspondence 235f., 391, 393, 397
- critical theory 82, 117
  
- deconstruction 4, 82, 112, 125, 243
- deduction 29f., 234
- deontology, deontontological 269–271, 273
- dependence 34, 36, 42, 297f., 301, 389
- description, descriptive 97, 140, 147, 188–190, 205, 215f., 252, 391, 395
- descriptive psychology 131
- disagreement 33, 40–42, 107f., 326, 329, 357–359, 363

- dispositions 204f., 207, 215, 217–219,  
266, 304, 385, 395, 398, 409f.
- 4e cognition 83, 143–153
- eidōs 157
- embodiment 148, 151, 203, 207f., 213,  
217, 325
- emergence 152, 206, 233f., 342
- empathy 121, 194f., 215, 218, 223, 268f.,  
279–282, 286, 313, 386
- enactive approach 143, 196, 216, 239
- envy 404
- epistemology, epistemological 4, 29, 34,  
60, 70, 86, 98, 111, 113, 115f., 144, 157,  
168, 183, 237f., 247f., 309, 311f., 315f.,  
319f., 372
- epoché 97, 177, 231, 239
- equality 35, 347
- ethics 53, 56f., 59, 263, 290–293, 349,  
369, 383, 401
- evidence 1, 6, 29–31, 39f., 51, 55, 86, 96,  
98, 113, 130, 134f., 182, 239, 313–315,  
317, 319f., 325–332, 356, 358, 365, 389
- exclusiveness 34, 36, 40–42
- existence 12–14, 42, 49f., 52f., 74, 84,  
98, 221, 225, 233, 247–257, 303f.,  
309–312, 319
- existentialism 117
- experience 12f., 15f., 18, 20f., 23f., 30,  
43, 57f., 71, 73, 84, 87–91, 95–108,  
112–114, 124, 132–135, 138, 149–151,  
187–190, 192f., 195–197, 204f., 211f.,  
216–218, 220, 222, 225f., 231–243,  
247–252, 254f., 257, 265f., 269,  
279–285, 297, 305, 309f., 312–320,  
324–326, 328–331, 333, 353, 361–367,  
384–389, 391–397
- experimental philosophy, X-phi 4
- externalism 149, 151
- first person, first-personal 13f., 20, 58, 68,  
87, 89, 95–105, 107, 188–190, 231, 235,  
237, 264, 278, 290, 292, 298, 302, 313,  
396, 401
- foundationalism 111–116, 118, 123, 126
- freedom 47, 49–54, 58–61, 73, 117, 119f.,  
126, 177, 181, 206f., 210f., 213, 227,  
263, 338, 340–342, 344–347
- fulfillment 132–135, 138f.
- German Idealism 111, 114, 124
- gestalt switch 99
- god 43, 98, 113, 118, 121, 124, 178–181,  
251, 337
- grief 64, 269
- group mind 272, 277f.
- habit 203–215, 217, 219, 221f., 393
- habituality 203f., 209–211
- hermeneutics 82, 111, 121f., 125
- hetero-phenomenology 97
- historicism 119
- history 1–4, 36, 43, 50, 82, 112, 117–119,  
121–123, 125, 182, 304, 338f., 341, 344,  
392
- hope 5, 13, 15, 25, 35, 64, 67, 101, 105,  
107f., 115, 187f., 264, 296, 332, 337,  
344, 366f., 384
- idealism 111, 115–118, 145f., 235, 238
- imagination 11, 108, 135f., 138, 143–151,  
153, 162, 212, 217, 247f., 250, 382
- impartiality 358
- imperialism 115, 123
- incommensurability 34
- independence 114, 296
- infallibilism 118
- intentionality 5f., 83, 96, 106, 139, 148,  
240, 272, 324f., 331, 333, 382–390,  
395
- intentionality, collective 83, 259, 262, 277,  
281, 285f., 302f.
- intercorporeality 215, 218, 223–225
- internalism, internalistic 148, 150
- intersubjective 152, 309, 318f.
- introspection 84, 95, 100–105, 107f., 123,  
187, 189f., 197f.
- introspectionism 84
- intuition 4, 129–141, 157–159, 164–172,  
300, 302, 369, 371, 373f., 378

- joint action 263, 289, 292, 300–303
- judgement 13, 31, 65, 67, 75, 114, 138, 358, 383
- judgment 36, 95–97, 99–101, 103–106, 135, 137f., 187, 231, 239, 241, 269, 284, 302, 304, 358, 381, 388, 394
- knowing-how 215, 217f.
- knowing-that 215, 217
- knowledge 29–32, 39, 59f., 70, 85f., 96, 113f., 116–119, 122, 124, 130, 132, 135, 138, 180f., 187, 189f., 204, 206, 215–221, 223, 225–227, 231, 233f., 238, 241, 243, 250f., 256, 264, 281, 293, 300, 304f., 309–313, 315f., 318–320, 343, 362, 365f., 392, 398, 401
- knowledge by acquaintance 215–217, 225, 227
- knowledge by description 215
- language 4, 11, 18–20, 33, 56, 71, 74, 82–84, 86, 90, 96, 122, 153, 215, 218, 220, 282, 286, 296, 311f., 315, 317–320, 328, 346, 383, 386f., 391, 393, 395, 397
- language-game 17, 31–33, 41, 69f.
- law 48, 54, 66, 114, 118, 120, 337f., 340–346, 351–356, 359–361, 363, 366f., 392
- lifeworld 2, 121, 126, 199, 227
- linguistic turn 4, 81–83, 85
- logic 12, 18f., 71, 73, 114, 118, 133, 136, 176, 181f., 206, 251, 338
- logical positivism 116
- mathematics 30, 65, 114, 181, 251
- maxim 48, 375f.
- meaning 3, 17, 31, 33, 39, 50, 52–54, 59, 64, 71, 85, 97, 125, 132f., 135f., 139, 147, 181, 208, 212f., 220, 223–225, 227, 232, 241, 243, 253f., 304, 309f., 312, 316, 343, 345f., 386, 389, 391f.
- memory 29, 39, 58, 192, 215, 217f., 221–223, 225f., 315
- mentalism 187
- mental state 54, 89, 101, 129, 131f., 140, 151, 188f., 191f., 194, 197, 223f., 253, 272, 279–285, 310, 312, 315, 317
- metaphor 11, 20, 68, 72–74, 193, 248, 394
- metaphysical 15f., 79, 83, 85f., 88, 120, 145, 203, 207, 234, 238, 264, 267, 312, 345
- metaphysics 4, 83, 86, 115, 144, 231, 233, 235, 243, 262, 345
- method 3–5, 63, 76, 80–82, 97–99, 111, 113–116, 119, 123f., 133, 145, 157, 167, 169f., 187–190, 224, 232f., 239, 354, 365, 375
- methodology 1, 77, 95, 98f., 101, 168f., 188, 243
- mind 57–58, 96f., 112–116, 123, 125f., 135, 149, 152, 187–193, 195–198, 203, 205–207, 209f., 213f., 216, 218f., 224–226, 231f., 234f., 255, 265, 277, 279, 309–312, 314, 317–319, 326, 329, 332, 357, 359f., 367, 381, 384, 387, 390
- motive 224, 295, 318, 339
- movement 47, 55f., 58–60, 73f., 212, 219–223, 225–227, 249f., 253, 284, 315, 346, 362
- naturalism 86f., 90, 111f., 121, 123, 143
- nature 30, 36, 44, 52, 87, 115, 118f., 124, 203, 206–208, 214, 235–239, 241, 392
- necessity 47, 52, 58, 117, 207, 313, 341
- neo-Kantianism 82, 111f., 115
- neo-Thomism 111f., 115
- neuroscience 188, 197, 199, 203
- neutrality 35, 98, 100, 122
- nonsense 27, 50, 85, 294
- normativity 265, 383, 385
- nous 157f., 161, 172
- objective 12–14, 40, 83, 85f., 106, 123, 181, 188, 192, 219, 231–233, 235–239, 243, 266, 290, 293, 346, 351–364, 366f., 385, 391f., 394
- obligation 20, 64, 173f., 176, 178–183, 263
- ontology, ontological 6, 12f., 15f., 20, 126, 191, 234–241, 254f., 261–263, 290f.,

- 298 f., 309, 311 f., 319, 344–346, 386, 389
- pain 11–29, 68, 99, 193, 195, 240, 271, 297, 311, 313 f., 316 f., 382, 390
- panpsychism 231–233, 235–237, 239, 243
- paradigm 1, 143, 146, 148, 153, 197, 216, 294
- perception 89, 123, 132 f., 138–140, 149, 152, 187, 196, 212, 215 f., 218–222, 225, 235, 240, 242, 247, 249, 255, 270, 279, 300, 304, 311, 313, 316 f., 325, 331, 385–389
- personhood 198, 261–267, 271, 273, 277 f., 285, 323 f., 327, 330 f., 396
- philosophy of language 4, 79, 83 f., 90
- philosophy of life 82
- philosophy of mind 4 f., 63, 68–70, 79, 83, 87, 90, 135, 143, 147 f., 150 f., 153, 196
- physicalism 234–238
- physics 29, 33, 114 f., 236–239, 361
- plurality 34, 36, 79, 136, 239, 265, 386
- possibility 15, 51, 68, 74, 88, 103, 114, 150 f., 196, 285
- postmodernism, postmodern 111, 117 f.
- post-structuralism, post-structuralist 82, 117
- practice 33–36, 40–44, 47, 54, 59, 98, 221–223, 227, 263, 317–319, 328, 331 f., 392
- pragmatism 112, 123
- pretence 309–312, 315–320
- proposition, propositional 5 f., 30 f., 34, 38, 41, 43, 49, 63, 65, 71, 73–76, 119, 129–132, 134, 140, 145, 159, 167, 173, 175–182, 204, 218, 237, 251
- psychologism 6, 79, 115, 145, 152
- psychology 52 f., 63–65, 69, 84, 97, 116, 122, 131, 144, 187 f., 199, 216, 221, 266–269, 285 f., 389, 396
- rationality 1, 31, 42, 101, 173–180, 183, 278, 325, 333, 383, 385, 387, 395
- realism 1, 4, 82, 146, 294
- metaphysical 131
- moral 369, 372–374
- scientific 143
- speculative 82
- reasonableness 205, 351 f., 356–358, 361 f., 364 f., 367
- reasonable, the 31, 40, 43, 351 f., 354–367
- reductionism 115, 247 f.
- reflection 47 f., 54, 86 f., 95–97, 100–108, 112, 197, 203, 208 f., 232, 323 f., 326 f., 329–332, 343–346, 388, 390
- reflective equilibrium 371
- relativism 29, 31–37, 39–45, 115, 391
- representation 6–8, 90, 96, 119, 131, 133, 138, 148 f., 150, 187, 196, 216–218, 232, 236, 239
- representationalism, representationalist 4, 145–147, 217, 223
- responsibility 96, 122, 197, 263 f., 267, 281 f., 287, 323 f., 330–332
- Royaumont meeting 84
- rule 72–74, 76, 118, 212, 223, 225, 256 f., 264, 314, 342
- schizophrenia 220
- science wars 122
- scientism 115, 143
- second person 302
- seeming 129, 140
- self-consciousness 88–90, 96, 107, 191, 193, 235
- self-transparency 214
- situatedness 217, 241, 247
- skepticism 43–45, 95, 118, 125, 177–179, 183, 187, 190, 311 f., 316
- skill 107, 208 f., 215, 217–219, 221 f., 226 f.
- social ontology 263, 267, 277 f., 283
- sociology 35
- spontaneity 206, 388
- stance taking 323–327, 329–332
- structuralism 82
- style 41, 64, 81, 86, 107, 115, 123, 226, 242, 252, 313
- subjective 12–14, 20 f., 88 f., 99, 149, 152 f., 190, 196, 219, 247 f., 255, 289 f., 293, 295, 351–356, 358, 360–362, 366 f., 383, 385, 387–390
- symmetry 33, 35 f., 39 f., 42, 122

- teleology 111
- temporality 3, 210f.
- testimony 29
- theory of action 247f., 253–255
- thought experiment 4, 85, 149, 241, 247f.
- tolerance 36, 42
- transcendental idealism 131, 146 f., 238,
- transparency 192, 226, 241
- truth 5 f., 49, 72, 85, 113, 117, 173 f.,  
178–182, 250, 296, 298, 344, 346, 359,  
382, 392, 398
- underdetermination 36, 42–44
- understandable, the 367
- value 33, 42, 54, 96, 193, 268–270, 298 f.,  
320, 337 f., 340, 344–347, 371, 381,  
385–388, 390–395, 397 f.
- Verstehen 121, 159, 167, 374
- virtue 32, 56 f., 59, 289 f., 295, 298–300,  
302, 337, 401, 404
- weakness of will (akrasia) 408
- well-being 289–303, 305 f.
- will 47–60, 181, 207–209, 227, 249, 252 f.,  
255 f.
- worldview 32 f., 115, 119 f., 123 f.
- x-phi 4



