AN ANTHOLOGY OF PHILOSOPHICAL STUDIES

VOLUME 8

Editor
Patricia Hanna
University of Utah
USA

Editorial Board

Gary Fuller
Central Michigan University
USA

Carol Nicholson
Rider University
USA

Donald Poochigian
University of North Dakota
USA

Andrew Ward
University of York
UK

Joseph Naimo
University of Notre Dame Australia
Australia

Athens Institute for Education and Research

2014
Board of Reviewers

Maria Adamos
Georgia Southern University
USA

Daniel Considine
Metropolitan State College in Denver, Colorado
USA.

Katherine Cooklin
Slippery Rock University of Pennsylvania
USA

Chrysoula Gitsoulis
City College of New York
USA

Keith Green
East Tennessee State University
USA

Dimitria Electra Gratzia
University of Akron
USA

Philip Matthews
University of Notre Dame Australia
Australia

Michael Matthis
Lamar University
USA

Mark McEvoy
Hofstra University
USA

Chris Onof
Birkbeck College
UK

John Thompson
Christopher Newport University
USA

Kiriake Xerohemona
Florida International University
USA
AN ANTHOLOGY OF PHILOSOPHICAL STUDIES
VOLUME 8

Edited by
Patricia Hanna

Athens Institute for Education and Research
2014
AN ANTHOLOGY OF PHILOSOPHICAL STUDIES VOLUME 8
Table of Contents

1. Introduction
   Patricia Hanna 1

Part A: History of Philosophy: Ancient, Early Modern, 19th and 20th Centuries
2. The Ancients, the Vulgar, and Hume's Skepticism
   Maria Adams 5
3. Russian Modernism or Mysticism? Vladimir Solovyev as Philosopher
   Trina Mamoon 15
4. Temporal Being and the Authentic Self
   Joseph Naimo 27
5. A Critique of Aristotle’s Politics
   Pritka Nehra 39
6. Kierkegaard and Moral Guilt: Can the Ethical Forgive Significant Ethical Failure?
   William O'Meara 49
7. Thaumazein in Ancient Greek Philosophy and Wonder in the Writings of Ludwig Wittgenstein
   Ilse Somavilla 61

Part B: Philosophy of Language
8. Semantic Intentionality and Intending to Act
   Dale Jacquette 75
9. Scope of Semantic Innocence
   Jaya Ray 87
10. Peirce's Theory of Continuity and the Vindication of Universals against Nominalism
    Paniel Reyes-Cárdenas 99

Part C: Philosophy of Science and Philosophy of Mind
11. Thinking about Mental States
    Sinem Elkatip Hatipoğlu 115
12. Naive Realism and the Explanatory Gap
    Takuya Niikawa 125
13. Bohm's Paradox and the Conscious Observer
    Donald Poochigian 137
14. Two Dogmas of Reductionism: On the Irreducibility of Self-Consciousness and the Impossibility of Neurophilosophy
    Joseph Thompson 149
15. Amendments to the Theory of Recognition
    Sander Wilkens 161

Part D: Value Theory
16. Paraesthetics: Irvine School of Aesthetic Theory and Criticism
    Ewa Bobrowska 173
17. The Standard of the Reasonable Person: An Avoidability Approach
    Michelle Ciurria 187
18. Plea Bargaining’s Moral Controversies
    David Heise 199
List of Contributors

Maria Magoula Adamos is an Associate Professor of Philosophy at Georgia Southern University in Statesboro, Georgia, USA. Her research interests include philosophy of emotions, philosophy of mind, ancient philosophy, forgiveness, modern philosophy, and, in particular, Hume’s account of identity. She has published in a variety of journals and editions.

Ewa Bobrowska is an Assistant Professor at the Academy of Fine Arts in Warsaw. Her research focuses on postmodern American aesthetics and philosophical interpretation of contemporary American art. She also conducts postdoctoral research at the University of California/Irvine under the supervision of prof. Francisco J. Ayala. Moreover, she is a video artist and a painter. Exhibitions of her paintings have accompanied three significant philosophical conference. Her book, Parateoria: Kalifornijska szkola z Irvine, was published by IBL PAN in 2013.

Michelle Ciurria is a Ph.D. Candidate at York University in Toronto Canada. Contact information is 4700 Keele Street, Toronto, Ontario CANADA, M3J 1P3, mciru@yorku.ca

Sinem Elkatip Hatipoğlu is an Assistant Professor of Philosophy at Istanbul Şehir University. Her research interests include philosophy of mind and metaphysics, with a particular interest in consciousness, self-consciousness, the concept of a self and personal identity. Her most recent publication “Consciousness and Peripheral Self-awareness” appeared in Organon F. She is currently working on the phenomenon of radical misrepresentation of target states by higher order states in higher order theories of consciousness.

David Heise is Associate Professor of Philosophy at Humboldt State University; his research interests include ethics, Asian and comparative philosophy, and philosophy for children. His work has appeared in such journals as Essays in Philosophy and Questions: Philosophy for Young People.

Dale Jacquette is ordentlicher Professor für Philosophie, Abteilung Logik und theoretische Philosophie (Senior Professorial Chair in Logic and Theoretical Philosophy), at Universität Bern, Switzerland. He is author of numerous articles on logic, metaphysics, and philosophy of mind, and has recently published Symbolic Logic, Philosophy of Mind: The Metaphysics of Consciousness, Ontology, Wittgenstein’s Thought in Transition, David Hume’s Critique of Infinity, and Logic and How it Gets That Way. He has edited the Cambridge Companion to Brentano, the Blackwell Companion to Philosophical Logic, and for North-Holland (Elsevier) the volume on Philosophy of Logic in the Handbook of the Philosophy of Science series. His latest book, forthcoming from Springer Verlag, is Alexius Meinong: The Shepherd of Non-Being.
Trina R. Mamoon is Associate Professor of Russian and Foreign Languages at the University of Alaska Fairbanks, where she has been teaching since 1998. Her research interests include Russian nineteenth and twentieth-century literature, contemporary Russian culture, and Russian and European cinema. She has published articles focusing on Russian painting, classical music and opera, and the Chechen conflict. Currently, she is doing research on Russian women writers during wartime.

Joseph Naimo is a senior lecturer teaching in Philosophy and Ethics at the University of Notre Dame Australia, Fremantle, Western Australia. Regarded as a Process Philosopher whose active areas of research broadly include Consciousness studies, Metaphysics and Ethics. Areas of interest: Philosophy of Science, Comparative Religious Studies, and Continental Philosophy. Contact Details: School of Philosophy and Theology, University of Notre Dame Australia 18 Mouat Street, Fremantle, Western Australia 6959 Email: joe.naimo@nd.edu.au

Pritika Nehra is a Research Scholar in the Department of Humanities & Social Sciences in Indian Institute of Technology (I.I.T) Delhi, India. Her doctoral research centers on exploring the concept of Politics. Her Research interests include Political Philosophy, Ethics and Aesthetics.

Takuya Niikawa is a PhD candidate at Hokkaido University in Japan. His research centers on the question of how we should characterize perceptual consciousness. His most recent work focuses on the viability of naïve realism. His master thesis defends and develops John Campbell’s relational view of perceptual consciousness. Other research interests include free will and perceptual epistemology.

William O’Meara is a Professor of Philosophy at James Madison University. He has published a number of papers on the Ethics of George Herbert Mead, Alan Gewirth, Karl Marx, Freud, and Kierkegaard and also several papers on the Philosophy of Religion of Alfred North Whitehead, Karl Marx, and John Hick.

Donald V. Poochigian is Professor of Philosophy in the Department of Philosophy and Religion at the University of North Dakota, Grand Forks, North Dakota, USA. His current research interests focus on, but are not limited to, identity theory and set theory.

Jaya Ray is an Assistant Professor of Philosophy at Lakshmibai College, University of Delhi, India. Her research interests include philosophy of language, philosophy of mind and logic. Her most recent work focuses on belief as a central concern in contemporary philosophy of language and mind.

Paniel Reyes-Cárdenas completed his MPhil and PhD studies at the University of Sheffield, United Kingdom. He worked on Peirce's pragmatism and the metaphysics of Scholastic Realism. He has published articles on the
Philosophy of Mathematics, Metaphysics, Medieval Philosophy, Philosophy of Religion, Hegel, and Kierkegaard. He is founder of the Mexican Society for Metaphysics and Philosophy of Science.

**Ilse Somavilla** is a Post-Doctoral Researcher and editor; her research interests center on Wittgenstein, Ancient Greek philosophy, ethics and aesthetics. She has edited manuscripts and letters of Wittgenstein, e.g. *Denkbewegungen. Tagebücher 1930-1932/1936-1937; Licht und Schatten; Wittgenstein – Engelmann*. Besides, she has written numerous papers on Wittgenstein’s philosophy.

**Joseph Thompson** is Associate Professor of Philosophy & Humanities at the University of Alaska Fairbanks, where he teaches courses in the history of philosophy, ethics, aesthetics, and the arts as humanities. He specializes in Nietzsche and German philosophy, with special interests in religion and art; he has published articles in comparative religion, ethical theory, and applied ethics. Current research focuses on the intersection of philosophy and cosmology with religion, myth, literature and the arts.

**Sander Wilkens** teaches at the Technische Universität Berlin (Habilitation 2011). His research interests concern metaphysics and the conditions of human consciousness from a modern view, including collective consciousness. The analysis of consciousness relies upon the logic of the relationships of the faculties, and the framework also reaches out to any form of opposition and generalization. The Habilitationsschrift (*Ordination des Bewusstseins und Metaphysik*, forthcoming 2014) implies a succinct investigation of projection and its history.
CHAPTER ONE

Introduction

Patricia Hanna

This volume is a collection of papers selected from those presented at the 8th International Conference on Philosophy sponsored by the Athens Institute for Research and Education (ATINER), held in Athens, Greece at the St. George Lycabettus Hotel, from 27—30 May, 2013.

This conference provides a singular opportunity for philosophers from all over the world to meet and share ideas with the aim of expanding the understanding of our discipline. Over the course of the conference thirty-eight papers were presented. The seventeen papers in this volume were selected for inclusion after a process of blind-review. The volume is organized along traditional lines. This should not, however, mislead a reader into supposing that the topics or approaches to problems fall neatly into traditional categories.

The papers chosen for inclusion give some sense of the variety of topics addressed at the conference. However, it would be impossible in an edited volume to ensure coverage of the full extent of diversity of the subject matter and approaches brought to the conference itself by the participants, some of whom could not travel to one another's home countries without enormous difficulty.

Since its inception in 2006, the conference has matured, reaching what might be seen as adolescence. Part of this maturity is reflected in the nature of the proceedings. We now have a group of dedicated philosophers who are committed to raising the standards of this publication; as a result of their work, we are now able to ensure that each submission is blind-reviewed by at least 2 readers, as well as the editor and/or a member of the Editorial Board. I would like to take this opportunity to thank them for their extraordinary work: without them, nothing would be possible; with them, we may reach adulthood!
Part A:
History of Philosophy:
Ancient, Early Modern,
19th and 20th Centuries
CHAPTER TWO

The Ancients, the Vulgar, and Hume's Skepticism

Maria Magoula Adamos, Georgia Southern University, USA

Section III of part IV of Book I of Hume's Treatise entitled “Of the ancient philosophy” has been virtually ignored by most Hume scholars. Although philosophers seem to concentrate on sections II and VI of part IV and pay little or no attention to section III, the latter section is paramount in showing how serious Hume's skepticism is, and how Hume's philosophy, contrary to his intention, is far removed from "the sentiments of the vulgar". In this paper I shall first explore Hume's view on ancient philosophy as it is presented in section III, and I shall particularly focus on his discussion of identity and simplicity of bodies. Second, I shall argue that Hume's account of identity and simplicity in terms of qualities is at best unsatisfactory. Finally, I shall try to show that Hume's advice to hold a "moderate" skepticism cannot be taken seriously. On the contrary, Hume seems to hold an "extravagant" skepticism, since he claims that there is a contradiction between our most fundamental natural beliefs, as well as between our natural beliefs and philosophical reasoning.

Keywords: Hume, skepticism, identity, ancient philosophy, substance

Hume's Account of Identity

The section “Of the Ancient Philosophy” is just another application of Hume's naturalistic approach to the birth of ancient philosophical theories. According to Hume, our ideas of bodies are nothing but collections "form'd by the mind of the ideas of the several distinct sensible qualities, of which objects are compos'd, and which we find to have a constant union with each other"(Hume, p. 219). However, in our everyday experience we regard, mistakenly, these distinct sensible qualities as "One thing, and as continuing the Same under very considerable alterations"(Hume, p. 219). We therefore attribute (falsely) to the "acknowledged composition" of perceptions simplicity and to the variation of them identity. But this is a contradiction, Hume tells us, since on the one hand our senses perceive totally distinct and different qualities, and on the other hand, we believe that the combination of these discrete parts possesses a unity and simplicity which endure over time.
Although according to Hume the ancients’ ascriptions of substance and prime matter are capricious and false, they, too, arise from fundamental principles of human nature, and as such they are worthy of study: “…I am persuaded there might be several useful discoveries made from a criticism of the fictions of the antient philosophy, concerning substances…which however unreasonable and capricious, have a very intimate connexion with the principles of human nature” (Hume, p. 219).

Why do we “almost universally fall into such evident contradictions?” (Hume, p. 219) In order to answer this question Hume discusses our idea of “identity of bodies”. According to Hume, when ideas of the discrete yet successive qualities of objects are united together by a very close relation, then the mind is “fooled”, so to speak, and considers the succession of different and separate qualities as just one, “continuous” thing:

…[T]he mind, in looking along the succession, must be carry'd from one part of it to another by an easy transition, and will no more perceive the change, than if it contemplated the same unchangeable object. This easy transition is the effect or rather the essence of relation; and the imagination readily takes one idea for another where their influence on the mind is similar; hence it proceeds that any such succession of related qualities is readily consider'd as one, continu'ed object, existing without any variation. (Hume, p. 220) [My italics].

Here Hume uses his typical associationist approach. That is, the uninterrupted progress of ideas deceives the mind and for that reason it ascribes an identity to the changeable “succession of connected qualities”. So, when the object is observed continuously through a succession of small changes, there is an easy transition from one idea to another and we believe that we have the same (self-identical) object, which endures over time.

Let us assume for the sake of the argument that Hume is right concerning the easy transition of ideas when they are in a close relation. What happens when the relation between the ideas is not close any more? In such a case, says Hume, although small changes over time may go unnoticed, if we observe the object in two different periods of time then the changes become evident and the mind becomes aware of them: “the variations, which were insensible when they arose gradually, do now appear of consequence, and seem entirely to destroy the identity” (Hume, p. 220). This is where the contradiction starts: on the one hand, the mind, viewing the changes of the object, is reluctant to ascribe identity to it; on the other hand, it has a strong propensity to attribute identity to the object despite the changes it has observed. In order to solve the

---

1In section III Hume is also criticizing the ancient philosophy’s ideas of “substantial forms”, “accidents”, “occult qualities”, etc. However, given that all these are “refinements” of the ancient idea of substance and/or prime matter, and since Hume is also objecting to the ancient idea of substance and prime matter, I am focusing the present discussion on these.
contradiction imagination (our *deus ex machina* "feigns" something "unknown and invisible which continues the same under all these variations; and this unintelligible something it calls a *substance, or original and first matter*" (Hume, p. 220).

So, according to Hume, the vulgar is in contradiction when she ascribes identity to objects. If this is the case, I do not see what is wrong with the ancient philosophical system, since to all appearances it seems to actually salvage the vulgar system from the contradiction by creating the notions of *substance* and *prime matter*. For, according to Hume this is exactly what the philosophical system *ought* to be doing; namely, to "approach nearer to the sentiments of the vulgar" (Hume, p. 222). Moreover, if the identity claim of the vulgar is unavoidable as Hume seems to suggest, then it follows that the postulation of first matter by the ancients may also be unavoidable.

But perhaps Hume’s argument of identity would fare better if we adopt his (presumably correct) theory of the mind. If Hume is right that only perceptions of sensible qualities really exist, the only thing that the vulgar (and the ancient philosopher) is truly aware of is her own perception of qualities --not the object that has such qualities. As we have seen, when the vulgar does not observe any changes in the qualities [of the object] --when there is an easy transition of the mind -- she easily ascribes identity to it. Hume’s theory of perception can easily accommodate this.

A much more important question is what happens when the vulgar recognizes the changes in the qualities of the object? Does she still maintain that the object has the same qualities in order to ascribe identity to it? Or does she posit a substance or prime matter in order to maintain that the qualities are the same? That is, when the vulgar realizes that the qualities of the object have changed, doesn't she recognize that she has made a faulty judgment?

We have to notice here that Hume's argument of identity in section “Of the Ancient Philosophy” is given in terms of qualities, and not in terms of constancy and continued existence. Therefore, even if we adopt Hume’s theory of perception, the different qualities that the object acquires through time do destroy identity and it seems that nothing can salvage the vulgar’s identity claim. What is more, even if we postulate a 'substance' or 'prime matter' in this case, we are still having the same problem. For instance, if the qualities Q (of an “object” P), say an oak tree, are observed in two different and distant periods of time t and t’, the qualities Q [of P] at t would be completely different from the qualities Q’ at time t’. In this case, the vulgar would have to admit that P at t becomes P’ at t’. (Assuming again that Hume's theory about the independent existence of perceptions is true). If Hume is right that only perceptions of qualities exist, then, *even if the vulgar “feigns” a notion of substance, she cannot salvage her claim that the qualities remained the same over the two time periods* (t and t’). The qualities of P have changed and it seems nothing can make them identical. But then, since the postulation of substance cannot salvage the claim that P and P’ are qualitatively identical, our inclination to ascribe identity cannot explain *why* we (or the ancients) should feign a substance. It seems that unless we already have a notion of substance in
which all these different qualities inhere, our identity claim concerning different qualities does not really work. In other words, it seems that unless we have *already* a notion of substance, which enables us to say that the object remains the same despite the change of its qualities, the postulation of substance, *after* we notice that the qualities are different, does not help our identity claim. For such a postulation will not make us think that the qualities we had previously perceived as different are now the same. In order for our identity claim to work, we need to first form a notion of *simplicity* of substance so that we can form a notion of an object, and then will we be able to ascribe identity to it. This will enable us to say that the object we observed at time t is the same as the object that we observed at time t'.

If I am right, then Hume's explanation of how we, the vulgar, ascribe identity to objects cannot stand. That is, even if we accept his theory of perceptions, Hume's analysis is unsatisfactory, since it does not explain how our postulation of substance or prime matter solves the contradictions Hume accuses us of. This in turn shows two things: 1) Hume is wrong in his criticism of ancient philosophical theories' postulation of substance, since they, unlike Hume, are able to “remain close to the sentiments of the vulgar.” 2) The vulgar is not only left with contradictory beliefs, but also with desperation and distress; for, even if she adopts the judicious philosophical standpoint (such as Hume's) she will never be able to get rid of or at least explain her erroneous judgments. Indeed, Hume's unsatisfactory analysis leads to an extravagant skepticism, since it shows that not only is there no justification for our natural beliefs, but also that we are not in a position to give an explanation of them, even if we adopt Hume's judicious philosophical standpoint.

**Hume’s Account of Simplicity**

Is Hume's discussion of “simplicity of substances” perhaps more convincing? Hume uses an associationist approach here as well. When the mind observes an “object” whose co-existent parts are closely related to one another by a "strong relation", it considers the object as *one, single* thing: "The connection of parts in the compound object has almost the same effect, and so unites the object within itself, that the fancy feels not the transition in passing from one part to another. Hence, the colour, taste, figure, solidity, and other qualities, combin'd in a peach or melon, are conceiv'd to form one thing" (Hume, p. 221). However, here, unlike the case of identity, the mind is not aware of the error --at least within the vulgar standpoint:

---

1 Terence Penelhum in his paper "Hume on Personal Identity" also argues that Hume's account of identity fails to capture the way we, the vulgar, think. Yet, Penelhum fails to notice that whether we are in error or not in our identity descriptions depends on the ontology that we subscribe to.
Whenever it [the mind] views the object in another light, it finds that all these qualities are different, and distinguishable, and separable from each other; which view of things being destructive of its primary and more natural notions, obliges the imagination to feign an unknown something or original substance and matter as a principle of union of cohesion among these qualities and as what may give the compound object a title to be call'd one thing, notwithstanding its diversity and composition. (Hume, p. 221)

Let us assume that we are able to observe the world in "another light", as Hume suggests. Then, again, we are faced with contradictions. If we adopt the vulgar standpoint, we see simple things whose parts form a united whole. However, if we adopt the philosophical standpoint, then we realize that the object is made of several parts, which are distinct and loose. In order to free itself from the contradictions, imagination feigns [again] "an unknown something, or original substance and matter, as a principle of union or cohesion among these qualities, and as what may give the compound object a title to be call'd one thing, notwithstanding its diversity and composition" (Hume, p. 221). ¹ So, according to Hume, the mind postulates an “original” substance in order to salvage itself from the contradictions of perceiving the separate existences of the different qualities of the object, and its simplicity simultaneously.

Unfortunately this argument will be shown to be problematic as well. Let us agree with Hume (for the sake of argument) that sensible qualities are indeed separate existences. Suppose that the color of an apple and the sweet taste of the apple are distinct and separate existences in the sense that at some time in the future the color may remain the same, while the sweet taste disappears and vice versa. However, this does not show that the red color of the apple (at that moment) can exist by itself, separate from the taste or the other qualities of the apple. Indeed, in section VII of part I even Hume seems to deny the possibility of distinct and separate existence of qualities. In this section Hume seems to claim that color and shape are just distinctions of reason:

... [W]hen a globe of white marble is presented, we receive only the impression of a white colour dispos'd in a certain form, nor are we able to separate and distinguish the colour from the form. But observing afterwards a globe of black marble and a cube of white

¹If Hume thinks that viewing the object in another light is only possible within the philosophical standpoint, then he does not show nor does he explain how we, the vulgar, form the idea of a single, whole, thing. From what he says in this passage it seems that what comes naturally to us is the idea of oneness and simplicity of the object and only upon reflection are we able to see that we are mistaken in thinking that the object we observe forms a unified whole. But, one would expect Hume to say the opposite; namely, that we first perceive the distinct and separable qualities of the object, and, then, because the mind becomes uneasy and has the propensity to ascribe oneness and wholeness to the object the imagination feigns a notion of substance in which all the qualities inhere.
and comparing them with our former object, we find two separate resemblances, in what formerly seem'd, and really is, perfectly inseparable. After a little more practice of this kind, we begin to distinguish the figure from the colour by a *distinction of reason*; that is, we consider the figure and colour together, since they are in effect the same and indistinguishable; but still view them in different aspects, according to the resemblances, of which they are susceptible. When we wou'd consider only the figure of the glove of white marble, we form in reality an idea both of the figure and colour, but tacitly carry our eye to its resemblance with the globe of black marble: And in the same manner when we wou'd consider its colour only, we turn our view to its resemblance with the cube of white marble (Hume, p. 25).

Here it seems that the qualities of the apple, like those of the marble, cannot be distinct and separate existences -- in the sense that they can exist without each other-- after all. For, the color of the apple and the taste of the apple are only distinctions of reason, and as such they cannot be distinct separate existences. But, if this is the case, then Hume's analysis of simplicity of substance is, again, unsatisfactory. For, if Hume is not able to show that perceptions of qualities have separate existences and are therefore distinguishable, then it seems that we, the vulgar, are not committing errors when we perceive the object as a unified whole. This in turn, can explain how we ascribe identity to an object over time. That is, we first form the notion of simplicity of substance where all these qualities inhere, and then we are able to say that an object remains the same even if its qualities have changed.

In his quest to answer why we fall to contradictions regarding the simplicity of substances, Hume tells us that the answer lies in the habit of the imagination. Surprisingly, he equates our habit of inferring from causes to effects to that of inferring a substance or accidental matter: "the same habit which makes us infer a connexion betwixt cause and effect makes us here infer a dependence of every quality on the unknown substance"(Hume, p. 222). But, if this is the case, then, one wonders, why Hume makes such a big fuss about ancient philosophy, claiming that these philosophers are worse than children and poets? If the same habit, which makes us infer from causes to effects makes us ascribe a "dependence of every quality on the unknown substance", then the conclusion of the ancients comes naturally. Moreover, if the habit that makes us infer a "dependence of every quality on the unknown substance" is the same as that which makes us infer from causes to effects, then it follows that both are emerging from the same principles of the imagination. ¹

¹However, in the beginning of the section "Of the Modern Philosophy", where Hume distinguishes between the two principles of imagination, he takes that these habits (i.e. the habit of inferring from cause to effect and that of inferring a dependence of every quality on substance) are entirely different, coming from different "principles of imagination".
Likewise, if Hume holds that imagination is the only "judge" in all philosophical systems, then how can he justify his own philosophical theory? His answer rests on the opening paragraph of the section “Of the Modern Philosophy”. In a moment of self-criticism Hume confesses: "But here it may be objected, that the imagination, according to my own confession, being the ultimate judge of all systems of philosophy, I am unjust in blaming the antient philosophers for making use of that faculty, and allowing themselves to be entirely guided by it in their reasonings” (Hume, p. 225). Hume's answer to this objection is that there is a distinction between two principles of imagination:

In order to justify myself, I must distinguish in the imagination betwixt the principles which are permanent, irresistible, and universal; such as the customary transition from causes to effects, and from effects to causes: And the principles, which are changeable, weak, and irregular; such as those I have just now taken notice of [concerning substances, substantial forms, accidents and occult qualities]. The former are the foundation of all our thoughts and actions, so that upon their removal human nature must immediately perish and go to ruin. The latter are neither unavoidable to mankind, nor necessary, or so much as useful in the conduct of life. (Hume, p. 225).

Hence, our belief in causes according to Hume is universal and unavoidable, whereas the ancient belief in substance is neither useful nor necessary in our everyday experience.

Consequently, for Hume the first set of reasoning (i.e. our belief in causes) is completely unproblematic. However, the second set is not only problematic but also useless." Now, even if we grant that the ancients' postulation of substance and prime matter is totally unnecessary and problematic, then what Hume says about identity in section II of part IV of the Treatise, seems rather odd, for it is highly dubious that we, the vulgar, think the way that Hume describes. We, like the ancient philosophers, take it that there is something more in the world other than our perceptions of qualities. That is, we, the vulgar, think that the world consists in tangible, material objects, which remain the same over time. Consequently, the ancient “fiction” of substance belongs to the first kind of principles of imagination Hume mentions, because it seems to be "unavoidable, irresistible and universal".

---

1It is worth noting that for Hume the philosophical system is always dependent on the vulgar system, since it does not have any authority of its own. One would expect the philosophical system to abandon the ideas of identity and simplicity since there are not sufficient grounds for even holding these ideas. However, this is not the case. Of course, the reason for this, according to Hume, is that nature renders the task of abandoning the ideas of identity and simplicity impossible.

2This directly contradicts what Hume had claimed in the previous section, (i.e. our belief in substance and our belief in causes are stemming from the same principles of imagination).
Hume’s Skepticism

A true philosopher according to Hume, is characterized by a moderate skepticism. Hume's advice to a “true philosopher” is to escape first from false philosophy and confess that "we have no idea or power or agency, separate from the mind" concerning the necessary connections in nature. For, "what can be imagin'd more tormenting, than to seek with eagerness, what for ever flies us; and seek for it in a place, where 'tis impossible it can ever exist?" (Hume, p. 223). Hence, the philosopher should attain "true philosophy" by returning back to the situation of the vulgar and regard "all these disquisitions with indolence and indifference"(Hume, p. 223).

However, Hume seems to hold an “extravagant” skepticism, for as we have seen, according to him when our natural beliefs are subjected to a critical reflection they lose any kind of justification whatsoever. Therefore, Hume's advice to hold "moderate skepticism" cannot be taken seriously. Indeed, Hume's insufficient account of identity leads to the most "extravagant skepticism", and I do not see how Hume can escape the charge of contradiction. That is to say, if Hume, who is a “true philosopher” is not able to give a sufficient explanation of our natural belief in identity, then it seems that nothing can really lead us to the truth. But, this conclusion (again) exceeds the limits of "moderate skepticism".

Conclusion

In summary, this paper is an attempt to show that Hume's account of identity in terms of qualities is inadequate. Hume's argument concerning the "uselessness" of the ancient philosophy's postulation of substance and prime matter fails to convince us. If indeed Hume believes that true philosophy should be nearer to the "sentiments of the vulgar" it seems that ancient philosophy fits exactly with this characterization, since it allows for a more plausible account of identity than Hume's. Accordingly, Hume's call to hold a "moderate skepticism" cannot convince us, since Hume himself seems to hold an "extravagant" skepticism by presenting the fundamental contradiction between our natural beliefs, as well as between our natural beliefs and philosophical reasoning. Besides, Hume's insufficient explanation of identity and simplicity of substances does not leave any room for 'moderate' skepticism whatsoever.

References


CHAPTER THREE

Russian Modernism or Mysticism?
Vladimir Solovyev as Philosopher

Trina Mamoon, University of Alaska Fairbanks, USA

The place of Russian philosophers has always been problematic within the Western philosophical tradition: the two most highly acclaimed thinkers in Russia, Vladimir Solovyev and Nikolai Berdyaev, are primarily philosophers of religion, and so in the West they are not considered to be fully philosophers in their own right. Western philosophy and religion have been divided into two autonomous disciplines, while in Russia the fields are closely related with little demarcation between them. The interconnectedness between philosophy, religious thought, and literature in Russian culture may be seen in the writings of Tolstoy and Dostoevsky, both known not only for their literary masterpieces, but for the philosophical and ethical dimensions of their work and thought.

Widespread amongst Russian cultural theorists is the view that the period of Russian Modernism (1880s-1920s) produced several philosophers, chiefly Solovyev and Berdyaev. While both thinkers have written extensively on topics concerning metaphysics, eschatology, and ethics, their arguments and premises are fundamentally grounded in orthodox Christianity, a tendency shared by almost all thinkers of Russian Modernism. None of the Russian Modernist philosophers were able to make a comprehensive break with religion and mysticism, a prerequisite for modern philosophy in the Western post-Kantian sense of the word. Those thinkers who made no recourse to religion and in fact rejected it—Alexander Hertzen, Mikhail Bakunin, and Leon Trotsky, among others—were materialists whose work forms the core of Soviet Marxist-Leninist philosophy.

While Solovyev may be only a philosophe, and not a philosopher proper from the Western standpoint, I will argue that under a broader interpretation of what philosophy is, his work must be considered primarily philosophical and not just “mystical,” a designation which carries negative connotations in Anglo-American analytic philosophy, but which represents no objection from a Russian point of view.

Russia is a land of contradictions, enigma and extremes, and cannot therefore be measured or evaluated by ordinary considerations. As in other respects, so in the area of philosophy, Russia holds a unique if problematic
place. A brief look at the table of contents of a handbook of Russian philosophy would show why—from a Western philosophical tradition—Russian philosophy is not a straightforward, clear-cut academic discipline. Many of the philosophers listed are notable writers, religious thinkers, social activists, and revolutionaries, and only a handful are professional philosophers, i.e. academic philosophers whose only or primary concern is to study philosophy systematically as a body of knowledge. True to long-standing tradition, Russians do not perceive any discrepancies in considering Feodor Dostoevsky, Leo Tolstoy, Mikhail Bakunin, and Alexander Hertzen all to be philosophers in their own right, even though the first two are world class writers, Bakunin a revolutionary, and Hertzen a political activist and theoretician. Anticipating objections to the Russian philosophical tradition, Kelly (1998), in her opening remarks on an essay titled ‘Russian philosophy,’ observes that ‘Russian thought is best approached without fixed preconceptions about the nature and proper boundaries of philosophy.’ A similar assertion is made by Copleston (1986) who advises his readers ‘to take a broad view of the relevant field and not to worry much about distinctions between the history of philosophy, the history of ideas, and the history of social theory and religious thought.’

Unlike the Anglo-American tradition where philosophers are professional academics engaged in the systematic inquiry into the various branches of philosophy, Russian philosophers—or as they are more commonly referred to in Russian as mysliteli (thinkers)—blend religion, social activism, and philosophical thinking in their writings. In this respect Russians are much closer to those ancient Greek thinkers whose philosophy was informed by religion. In the contemporary Western philosophical canon, there is a marked distinction between religion and philosophy. The crux of the contention between Russian and Western philosophy lies in this very distinction. By this measure, Vladimir Solovyev, who is considered by Russians to be the Russian philosopher par excellence, would not be considered a philosopher proper, but a philosophe in the vein of Rousseau and Voltaire, or a mystical philosopher such as Simone Weil.

One of the reasons why Solovyev is considered by Russians to be their foremost philosopher and is held in such high esteem is the fact that he was the first of the Russian philosophers who undertook the methodical study of philosophy and produced a number of works concerning the different branches of philosophy: ethics, epistemology, aesthetics, and metaphysics, as well as publishing a treatise on the history of modern Western philosophy. His philosophical works include publications such as The Crisis of Western Philosophy (1873), The Philosophical Foundations of Integral Knowledge (1877), and Criticism of Abstract Principles (1878). In the mid-1850s when Solovyev was engaged in his philosophical pursuits, German philosophers were widely read and debated in Russian intellectual circles, and members of the Russian intelligentsia were familiar with the works of contemporary German philosophers such as Kant, Hegel, Schelling, Schopenhauer and
others. Solovyev may be best described as a mystical Kantian, who instead of just reinterpreting and reworking Kant’s philosophy actually added his own dimension to it. Seeking synthesis between philosophy and Christianity, Solovyev developed Kant’s theory of knowledge; to Kant’s two sources of knowledge, reason and experience, he added intuition. The very contribution that Solovyev made to philosophy, paradoxically, causes him not to be considered a philosopher proper in the Western canon, and he is therefore associated with that particular branch known as mystical philosophy. Expounding on Solovyev’s metaphysics, Copleston (1986) makes the following remark:

[To western philosophers who are representatives of the analytic current of philosophical thought this speculation (metaphysical and theosophical) is apt to seem fantastic. When Solovyev is discussing empiricism or rationalism, or criticizing Descartes, they see him as a philosopher, whether or not they agree with all that he says. But when he starts talking about the Absolute and Sophia and Godmanhood, they probably think that his thought belongs to another world.

At this point it would be relevant and useful to offer a brief overview of the Russian social and cultural scene of the nineteenth century, the period when Solovyev was producing his works. Copleston’s observation on Solovyev’s metaphysics cited above would hold true for most other non-Marxist Russian philosophers such as Nikolai Berdyaev, Lev Shestov, or Vasily Rozanov whose philosophical foundations were informed by a Christian conception of the world. What is of note here is that these religious thinkers were active around the time of Modernist activity in Russia, a dynamic period known for its rejection of religion and traditional values. Russian Modernism (1880s-1920s) was distinguished by the engendering of new ideas and new movements in the arts and literature, and is associated with the names of such illustrious cultural figures as Vasily Kandinsky, Marc Chagall, and many more. The poetics of the Modernists such as Vladimir Mayakovsky was to ‘épater le bourgeois’ (to shock the bourgeois) and they were instrumental in creating bold, new modes of thinking in the period just preceding the October Revolution of 1917. Russia, being the land of contradictions and discrepancies, alongside the radical Modernists also produced a movement that was deeply religious or mystical in character. Solovyev was an important figure who exerted a deep influence on a mystical group of poets and writers known as the Symbolists or the poets of the Silver Age. More will be said later about Solovyev’s impact on this movement.

One of the circumstances that solidified the ties between philosophy and religion and spirituality in Russia is the fact that since the 1820s Russian identity was defined in respect, or in opposition, to Western civilization. Russia’s identity as a nation started to be reshaped as the ‘Other’ of the West,
namely in opposition to its material progress and secular institutions. This major upheaval in self-identification, an identity crisis of sorts, emerged as a result of the publication of a series of letters—*Philosophical Letters*—by Pyotr Chaadaev. In the first of the letters, published in 1828, Chaadaev compared Russia, detrimentally, to the West. Chaadaev not only denounced serfdom, he also claimed that Russia had made no valuable contribution to the world, and anything that may be of value created by Russia, originated in the West. He even went further and extolled the role of the Catholic Church in advancing the idea of social progress in the West (Copleston, 1986). Even before Chaadaev’s open letter, the Russian intelligentsia was divided into two social-political camps: Slavophiles and Westernizers. As the name suggests, the proponents of the Slavophile camp saw the future of Russia in indigenous Russian values and adhered to an Orthodox Christian conception of the world. The Westernizers, on the other hand, were mostly atheists, and believed that Russia should found its institutions on the progressive, democratic principles espoused in the rest of Europe. Distinguishing between the two camps, Chamberlain (2004) observes

So what was the difference between the Westernizers and the Slavophiles in the end? It was in their underlying philosophy of ethics in relation to knowledge. The Slavophiles were religious conservatives, the Westernizers atheist progressives, which put them in different philosophical camps with regard to science and reason. The Slavophiles, perfectly represented by Khomiakov in this respect, were skeptical of the civilizing power of reason.

Slavophiles espoused ‘Hegel’s model of progress, which started with naïve community,’ while the Westernizers ‘looked forward to that social complexity’ that Hegel later moved on to (Chamberlain, 2004). Solovyev belonged to the Slavophile camp that espoused the view that politics, law, and all other aspects of civic life should be governed by religion. And like so many Russian philosophers who came before 1917, he was keen on linking philosophy with religion, and reason with intuition. Characterizing Russian philosophical thought and its connections to religion, Lossky (1951) observes that unlike the hard sciences, philosophy represents its national character, and therefore one may speak about the national peculiarities of German, French, English, American, and Russian philosophy. And accordingly, the subjects of study are based on each nation’s interests and experience such as the sensuous, the practical, or religious experience. Discussing the ideal of integral knowledge in Russian philosophy, Lossky explains:

The whole truth is revealed to the whole man. ... It is only through combining all his spiritual powers—sense experience, rational thought, aesthetic perception, moral experience and religious
contemplation—that man begins to apprehend the world’s true being and grasp the superrational truths about God.

Lossky’s observation would seem not only to lend support to the Russian philosophical viewpoint that integral knowledge is as valid as empirical knowledge, but that integral knowledge is more complete, as it contains the ‘whole truth.’ In this connection Zouboff (1944) notes:

... while there is no doubt that in its technically philosophical aspect Slavophilism was wholly based on Hegel and still more fully followed Schelling, it should not be forgotten that for the Slavophiles philosophy, especially metaphysics and epistemology, was subsidiary to their major theme of the Church and the State viewed in the light of the Christian conception of history.

This statement helps shed light on the Slavophile mindset, and consequently on the mindset of Russians in the late nineteenth century. It should be noted here however that this mindset or outlook is prevalent even today when there has been a revival of Russian nationalist sentiment in the new post-Soviet Russia. Russians generally tend to consider religion and theology to be legitimate branches of philosophy. Given this position, it may not be too problematic to accept Solovyev’s stance on religion—specifically his belief in Christian eschatology—as a philosophical premise rather than theology.

Given the Zeitgeist of Russia of the nineteenth-century, it is not difficult to see how Solovyev, initially a student of mathematics and physics, became inspired by religion. Solovyev was born in 1853, a time when Russian society was undergoing major changes in all the different spheres. On the political front the notable Populist movement—the Narodniki—was strong at that time, along with the Emancipation of the Serfs that was decreed in 1861 by Tsar Alexander II. On the cultural front prominent writers like Tolstoy and Dostoevsky were at the height of their creative genius in the 1860s; masterpieces such as *Anna Karenina* and *Crime and Punishment* were produced in that decade. Solovyev’s worldview was not only shaped by this turbulent yet productive period in Russian history, his family background also played a significant role in his religious formation. Born into an academic family—his father was a renowned historian—he was deeply influenced by his grandfather who was a priest. While Solovyev experimented with an atheist phase for a brief period of time, from early childhood he had a mystical bent. In a poem written shortly before his death he describes his first mystical experience that he had during a church service at the age of nine. He believed himself to have had a vision of Sophia, the ‘Eternal Feminine’ and the principle of divine wisdom in Eastern Orthodoxy. This experience was to have a lasting impact on his life. Solovyev’s contribution to Eastern Orthodoxy was his concept of the Divine Sophia as the unifying principle of God, humankind, and the universe.
Even though Solovyev is best known and respected in his native land for his religious and mystical writings, he actually produced works in philosophy. Interestingly enough, his philosophical works come at the very beginning of his career as a philosopher and towards the very end of his life. According to Lossky (1951), while the beginnings of independent philosophical thinking in Russia can be traced back to Ivan Kireyevsky (1806-1856) and Alexei Khomiakov (1804-1860), it was Solovyev who was ‘the first to create a system of Christian philosophy in the spirit of Kireyevsky’s and Khomiakov’s ideas.’ In Russia positivism was at its height in the 1870s, and Solovyev’s first philosophical work *The Crisis of Western Philosophy (Against the Positivists)* published in 1874, as the title would suggest, is a ‘dense critique … of [the] entire philosophy of Western rationalism from John Scotus Erigena onwards’ (Jakim, 1996). In this seminal work Solovyev divides the different stages of the development of Western philosophy into three major currents or phases: 1. faith as the prevailing authority, 2. faith and reason as equally dominant, and 3. reason as the prevailing authority (Jakim, 1996). According to Solovyev, all three currents are one-sided and lack synthesis. The idea of synthesis and unity between reason and faith, intuition and empirical knowledge, and unity between humankind, the universe, and the divine is a thread that flows through Solovyev’s entire body of work. As he was to argue along the same lines in his later works, Solovyev tried to make a case against Western philosophy—the positivists—claiming that it had reached an impasse. Solovyev contends that contemporary Western philosophy—pure rationalism and pure empiricism—is impoverished as faith plays no part in this canon. Solovyev was distrustful of the scientific knowledge of the positivists that disapproved of faith and intuition as valid sources of knowledge.

In *Crisis*, offering both an overview and an in-depth analysis of modern Western philosophy, Solovyev demonstrates that just as human beings consist of substances and qualities that range from lower (chemical and organic substances) to higher forms (consciousness and spirituality), so is philosophy built upon various stages of development. According to Solovyev the higher or later stages of philosophy are built upon the earlier stages that lacked synthesis and inclusiveness; later stages of the development of philosophy have strived towards synthesis. In *Crisis* Solovyev took upon himself the task of explaining to his Russian audience the dispute between the rationalists and the empiricists.

Following the Western philosophical tradition, Solovyev divided modern Western philosophy into pre-Kantian and post-Kantian, and argues that starting from Descartes, each philosopher or philosophical school was more developed than the previous one. Therefore, not surprisingly, Solovyev denounces Descartes for being the foremost proponent of rationalism—a school of thought that he strongly disapproved of—because of its stance on spiritual intuition as a source of knowledge. As for Spinoza’s monism, while Solovyev finds that it has progressed from Descartes’ philosophy, he considered it to be inadequate to serve as a viable philosophical foundation. The Western philosopher whose theory Solovyev found to be closest to his own was the German rationalist
Leibniz. Leibniz’s philosophical works treated the questions of free will, sin, and good and evil, questions that were of deep interest to Solovyev himself. Moreover, Leibniz was a theist who maintained that God chose the best possible universe, and preestablished harmony between the mind and body. Leibniz is one philosopher who permitted the consideration of God in his work, and therefore, Solovyev does not outright refute Leibniz’s theory. The empiricists, or the Anglo-Scottish philosophers, as Solovyev sometimes refers to Bacon, Hobbes, Locke, Berkeley, and Hume, on the other hand, were the subject of intense scrutiny and examination by Solovyev in this work. The positivists, as Solovyev scornfully referred to them, collectively represented for him the ills of the ‘new philosophy.’

Discussing Bacon’s work, Solovyev shows his intense distaste for the ‘new brand’ of philosophy: empiricism. Solovyev states that Bacon is only important in so far as he happens to be the founder of the new school of empiricism; apart from that, Bacon’s work, according to Solovyev, lacks import. He dismisses Bacon’s work as ‘vulgar,’ claiming that it lacks philosophical character. Solovyev sums up Bacon’s views in a few short sentences, disparaging him for wanting ‘to free the mind from deceitful suppositions or preconceived notions.’

Turning to Hobbes’ contribution to philosophy, Solovyev once again uses the word ‘vulgar’ to evaluate it, since for Hobbes everything is comprised of physical substance. While Solovyev considers Locke’s philosophy to be ‘interesna’ (interesting), his own interpretation of Locke is an interesting one too: he calls Locke a subjective idealist, pairing him with Berkeley and not with Bacon and Hobbes. In Solovyev’s view, Locke deviates from the objective realism of his predecessors and ‘twists’ in the opposite direction. For Solovyev, Locke’s theory that the human mind possesses no innate knowledge and is a tabula rasa implies that all knowledge about the external world is subjective. Therefore Solovyev claims that by Locke’s view ‘all matter boils down to subjective elements, and exists in our imagination and does not exist on its own,’ which would make Locke a subjective idealist rather than objective realist. Since Hume was a skeptic and an atheist, or a polytheist at best, he is denigrated as having had profoundly negative implications for metaphysics, just as Kant had read Hume. Summing up Hume’s theory of relationality and causality, Solovyev remarks that Hume reduced the objective world into the chance sequencing of unrelated ideas, and truth to the unconditional unknown x. According to this view no metaphysical knowledge would be possible.

For Solovyev Hegel’s philosophy of ideas and concepts based on logic and objective truth embodies the inadequacy of Western rationalist philosophy, its inability to compete with religion as far as the question of theory and practice was concerned. Solovyev argues that the domain of philosophy, by its very nature, is limited only to theorizing and lacks the capacity to affect change. Solovyev states that if a philosopher in his capacity as philosopher wished to bring change in real life through his theories of ideal orders and norms of
actuality, then theory would remain theory and would not yield any practical outcome. Solovyev asserts that philosophy cannot achieve a ‘double victory’ over the faith of people or over the social order that is founded on that faith. If faith is strong and the social order sound, then philosophical thinking is powerless to influence any change.

In Solovyev’s view, religion on the other hand—he names the Catholic Church—he can and does have the ability to influence the mind of an entire nation and thereby usher in practical changes in the form of social institutions. Given the fact that at the time when Solovyev wrote this treatise, his motherland Russia had an autocratic monarchy and consequently lacked strong secular institutions, it is perhaps not surprising that he would discount Hegel’s ideas about the practical side of philosophy. For Hegel scientific, social, political, and legal institutions manifested the practical aspect of philosophy. Western Europe, by the mid-nineteenth century, at the time when Solovyev was at the height of his career, had secular institutions firmly in place and did not require the authority of the church or religion to guarantee its citizens certain basic civil rights.

It is ironic that Solovyev perceives the inadequacy of Western philosophy due to its lack of practical application, when at the time he was writing his philosophical works, Marxist philosophy had become influential in revolutionary circles in Russia. Following Marx’s famous thesis that ‘the philosophers have only interpreted the world, in various ways; the point is to change it,’ radical Russian political movements were espousing and advancing the theories of Marx and Engels to change their world, their social order. Perhaps it was precisely because Solovyev saw the signs that positivist materialist philosophy—the very philosophy that he was so critical of—was rapidly gaining ground in Russia, and its influence was becoming widespread, capturing the hearts and minds of his fellow countrymen, that he chose not to recognize it. And perhaps that is why even when he discusses socialism and denounces it, he does not mention Marx or Engels by name. It was as if by not including these two philosophers who were succeeding in making real change in social and political institutions, Solovyev could halt the dissemination of this materialist philosophy in tsarist Russia.

Solovyev concludes his critique of Western philosophy by asserting that positivism as a philosophical school was untenable and unsound, as in Solovyev’s reading, positivists affirm the system of empirical sciences as the only true knowledge and negate all unconditional beginnings, religious and philosophical. This position, in Solovyev’s view, confirms the limitedness and harmfulness of positivism.

Solovyev’s renown as philosopher does not rest on his purely philosophical works, but is derived through his series of lectures on the Eastern Orthodox concept of ‘Godmanhood’: bogochelovechestvo. In his first lecture on ‘Godmanhood’, delivered in Moscow in 1876, which drew a large audience, Solovyev expounds his vision that the ultimate objective of history is the union of the divine beginning with humankind. The changes in social organizations
and the evolution of religious beliefs all represent the preparatory stages of this union of God and man (Zouboff, 1944). This idea was based upon his conviction that Russia and Eastern (Russian) Orthodoxy are historically destined to be the savior of all humanity. In his view the French Revolution, while having espoused the principles of equality, liberty, and brotherhood, did not accomplish its goals of equalizing people from a moral point of view. Solovyev grants that the French Revolution ‘established civil liberty,’ but given that there still existed ‘social inequality’ in France (in the late 19th century at the time his lectures were delivered), the ‘emancipation,’ in Solovyev’s opinion, took place only from ‘one dominating class to subjugation to another.’ Solovyev contends that true equality can only be achieved through the practice of a Christian social order that is based on ‘unconditional, supernatural and superhuman’ principles. The idea of the unconditional beginning lay at the core of Solovyev’s belief in religion, and the unconditional nature of it, according to him was what rendered religion its moral superiority over socialism, positivism, or other philosophical theories of social justice and civil liberties. Making a case in favor of a Christian social order, Solovyev (1944) lays out his conception of religion:

Religion, speaking generally and abstractly, is the connection of man and the world with the unconditional beginning, which is the focus of all that exists. It is evident if we admit the reality of this unconditional beginning, it must define all the interests and the whole content of human life; consciousness must depend upon it; and to it must be related all that is essential in what man does, learns, and creates.

Solovyev concludes his series of lectures with a call for accepting the Godman, Christ, not just externally, but internally, as he believes that this free acceptance will lead to the regeneration of a new, spiritual man, a spiritual humankind that will accept the law that was given ‘in the revelation of Christ.’ He firmly believed in the Second Coming of Christ when Christ the Godman would transform humankind into the state of ‘mangodhood’, a state that would elevate human beings to be God-like. This progression would thus usher in the final historical era, an era distinguished by universal peace and harmony, an era where there no longer will be sin, evil, or suffering.

The other fundamental concept that was connected with the idea of ‘Godmanhood’ was the idea of the Divine Sophia, the principle of the Eternal Feminine, a concept that he promoted and developed into a cult-like status. For Solovyev, Sophia represented the world soul and therefore was the integral link between God (the transcendental being) and human beings (nature). Explaining the concept of Sophia in relation to ‘Godmanhood’, Solovyev (1944) states:

Insofar as she receives unto herself the divine Logos and is determined by Him, the soul of the world is humanity—the divine
manhood of Christ—the body of Christ, or Sophia. Conceiving the
unitary divine beginning and binding by this unity the entire
multiplicity of beings, the soul of the world thereby gives the divine
beginning its complete actual realization in everything: ... God is
manifested in all creation as the living, active force, or as the Holy
Spirit.

Originally envisioned as a religious concept, the idea of the ‘Eternal
Feminine’ developed into a cultural phenomenon, as it exerted an enormous
influence over a new generation of poets—the Symbolists or poets of the Silver
Age—a mystical branch of Russian Modernism. The Russian Symbolist
movement which was comprised not only of poets, but also of novelists,
playwrights, artists, and philosophers is greatly indebted to Solovyev’s vision
of the Divine Sophia and the ‘Eternal Feminine.’ At the core of the worldview
of the Symbolists lay Solovyev’s mystical philosophy and it served as the
inspiration of these poets. The Beautiful Lady or the ‘Eternal Feminine,’
central themes in the poetry of the Symbolists, acquired a cult-like following
among the proponents of this school, as they considered the Divine Sophia to
be the embodiment of the world soul and the source of harmony on earth. It is
due to this very concept that many of the Symbolists hailed Solovyev as a
prophet. While Solovyev’s status as a philosopher may be a point of contention
and debate, his influence on Russian Modernist culture is profound and
uncontested.

Evaluated by the Western philosophical standard, Solovyev may not have
earned the title of philosopher proper, but he certainly was a philosopher of
religion by any measure. If he were to be judged by the yardstick of his own
standard, he was a philosopher in the sense that he preferred philosophers to be—individuals who are able to effect change and to have a lasting, practical
impact on a nation’s psyche. Solovyev succeeded in capturing the imagination
of his fellow citizens at a critical period in Russian history when Russia was
undergoing major change and transformation. Through his writings, both
philosophical and religious, he was able to have a profound influence on a
particular generation of cultural figures who in turn were able to offer a
different, idealistic vision of a future Russia on the eve of the October
Revolution of 1917. The status of Solovyev as a philosopher should not be
evaluated by the sole measure of the modern Western philosophical tradition.
His place and contribution as a philosopher should be assessed by a broader
matrix where it is permissible to bring in a mystical dimension to philosophy
and where there is room for intuition and integral knowledge to be
acknowledged as potentially valid sources of knowledge. If not a full-fledged
philosopher by the analytic model of philosophy, then Solovyev should be at
the very least granted the status of a philosopher-mystic who was, as Zouboff
(1944) writes, a ‘Christian pragmatist … fighting against the separation of the
practical from the ideal—from the Christian ideal.’ Solovyev’s legacy
continues, to this day, to serve as the inspiration for creativity and artistic
endeavors among Russians and Russophiles everywhere. His thought and works were instrumental in creating an understanding in modern Russia of philosophy as rigorous and systematic, and helped make philosophy more accessible to a broad audience. After Solovyev philosophy in Russia was no longer the exclusive province of professional, academic philosophers, it was made available, so to speak, in the public domain for debate and discussion.

Bibliography


Temporal Being and the Authentic Self

Joseph Naimo, University of Notre Dame Australia, Austrailia

The central issue raised here concerns whether Being as explored by Martin Heidegger in *Being and Time* is constituted spatiotemporally. As such this project has two interlinked objectives. One objective is to supply conceptually plausible answers to Heidegger’s unanswered questions regarding the temporality of Being, which he raised at the very end of *Being and Time*. In response I argue that each individual human being is constituted as a Space-Time-Event-Motion (STEM) containment-field embodied entity. Heidegger situates *Dasein* (human existence) in a temporal stream moving towards the nothingness of death but all the while separating being and time as two distinct phenomena rather than coexistent. I demonstrate why this was a crucial error in Heidegger’s thinking. The second objective deals with the sense of authentic being developed by Heidegger. The spatiotemporal nature of one’s life for Heidegger is understood from the standpoint of Being-in-the-world, as an engaged participant, coexistent with the world, so that contextually it is through this engagement in recognition of this facticity or thrownness that one may come to recognise one’s own authentic self.

Yet there is a tension between the individualistic sense of authentic self and the coexistent Being-in-the-world they-self (*Das Man*) that emerges from Heidegger’s analysis. Somewhat problematic then is the coexistent phenomenological recognition that one does not live in isolation and as such one may question what of the contingent, constraining and influencing factors that shape one’s sense of self particularly against the backdrop of self-other relations. This project is situated methodologically within Process Philosophy, and it is from this perspective that draws attention to the role of human agency in which individuals are spatiotemporally construed in terms of Space-Time-Event-Motion (STEM) entities.

**Introduction**

Heidegger, in Division 1 of *Being and Time*, posits two categories of Being: *Zuhandensein* (readiness-to-hand) and *Vorhandensein* (presence-at-hand). *Zuhandene* things are available things in the world assigned with human values and significances (Brandom, 2007: 214). In this way what we encounter
as being-in-the-world he calls ‘equipment’, so described since through our actions we are mainly engaged in tasks. ‘The readiness-to-hand of a piece of equipment (like a hammer) consists in its having certain significance’ based on its ‘appropriateness for various practical roles and its inappropriateness for others’ (Brandom, 2007: 217). For example, appropriate for hammering a nail into wood yet inappropriate for slicing bread. Vorhandene things do not have a worldly determination, as it seems they are objective, person-independent, causally interacting subjects, as in natural scientific inquiry. The term Vorhandene refers to the occurrence of things rather than through their use as tools or functional relations. Dasein stands as the third category ‘in whose structure the origins of the two thing-ish categories are to be found’ (Brandom, 2007: 214). As such the being of the ready-to-hand has in itself the character of having been assigned or referred (Brandom, 2007: 218). To reiterate each and every one of us is Dasein.

For Heidegger ‘moods’ reveal significant details about the fundamental structure of the world and also our way of Being-in-the-world in two subtle ways. ‘Moods assail us’, says Heidegger, disclosing that we are ‘thrown’ into a world not of our making, in that we find ourselves firstly in the world. Secondly, moods indicate something shared and essentially not always inner and private by tuning us into the world. Things in the world, like events and situations, therefore exude a quality that resonates with us as mood. Since for Heidegger Being-in-the-world reveals that we find ourselves in the world in a particular way such that we have a ‘there’, a meaningfully oriented situation in which to act and exist; hence we are disposed to things in ways that matter to us. Heidegger says that ‘Disposedness’ is an ‘attunement’ (Hubert, Dreyfus, & Wrathall, 2007:5). In other words an affective condition in so far as it discloses what matters to us when tuned in to things in the world. Feeling playful attunes one to the playfulness dimension of others. ‘Mood’, ontologically speaking, as a way-of-being, can be physiologically regarded as a transmittable dynamical medium construable in terms of communicable information (molecules/electromagnetic fields, usually visually enactive but most importantly viscerally as well) receivable by others (feedback/feedforward). This explains the inherent receptive capacity we humans exhibit that enables sharing, the dynamics of affective social tuning via certain moods (e.g. genuine smiles attract smiles; group dynamics at a musical concert, etc.).

Heidegger claims that our primordial experience is a unified experience of being in the world in an inseparable way. Heidegger was particularly critical of both Husserl and Descartes for their dichotomised view of mind and body. Which explains why Heidegger moved beyond Husserl’s articulation of the unfolding of ‘inner time’ (‘impression’, ‘retention’, ‘protention’) developing instead an understanding of ‘human beings as a ‘nexus’ of lived experience ... he recognises the priority of the ‘lived’ world [Lebenswelt] ... its three temporal dimensions ... ecstasies, in which we, as temporal beings, exist all at once’ (Ward, 2008:100) past, present and future. Indeed Heidegger recognises that ‘Being is essentially temporal’, in that ‘Being is always understood in terms of
Temporal Being and the Authentic Self

Time explained by its temporal structure’ (Blattner, 2006:14). Yet Heidegger’s claim requires demonstrating how the temporal structure of Being, indeed of each \textit{Dasein}, can be construed and understood. An endeavour remaining unanswered that Heidegger himself admits at the end of \textit{Being and Time}.

Arguably, then, one way to achieve this end requires a greater perspective shift than that taken by Heidegger whereby one’s existence is not simply observed as an entity occupying some volume of space for some duration of time but instead as being made of spacetime in a constitutional, compositional form. Heidegger, in \textit{Being and Time}, is at pains to articulate the rather fundamental element of temporality as integral to being. My contention is that the concept of STEM containment-field as an embodying concept provides the appropriate conceptual mechanism to further ground Heidegger’s endeavour.

The Spatiotemporal Fabric of Being

As established in previous work the concept of Space-Time-Event-Motion (STEM) provides a framework and perspective to facilitate inquiries into human nature construed as spatiotemporal beings consistent with current physical theories. Each able-bodied STEM (human being) is a moving space-time agent, an ongoing event in and of life whose own constitution is the subject within and of the environment (e.g. epigenetic process). Each STEM agent is construed as an ecosystem, operationally a self-organising multicellular integrated system of Being. Genetics alone provide an incomplete account of the mechanisation of human existence. Characteristically human bodies are largely reciprocating organisms and the subject of ongoing compositional change particularly epigenetic chemical change, cellular change, emotional and intellectual change, hence plasticity, to say the least that exist of and along a spatiotemporal continuum. Space and time are mutually interconnected. Indeed Leibniz’s principle of the \textit{identity of the indiscernibles} holds “that no two moments of time can be identical” (Smolin, 2013:214). Consequently, there cannot be two objects in the universe that are indistinguishable yet distinct (Smolin, 2013: 214). There is no absolute space so physical properties of bodies are relational. Lee Smolin argues that “according to the \textit{principle of the identity of indiscernibles}, our universe is one where every moment of time, and every place at every moment, is uniquely distinguishable from any other” (2013: 215).

From this perspective attention turns to Heidegger’s examination regarding what he described as three existential features of \textit{Dasein}: ‘existence’, ‘facticity’ and ‘fallenness’. \textit{Dasein}, Heidegger maintains, has no essence, yet it exists, since for Heidegger ‘existence precedes essence’. \textit{Dasein} has possibilities, whereby \textit{Existenz} refers to a kind of \textit{Being} relative to comporting oneself in a variety of ways that provides a sense of who one is through one’s continued actions. In one sense, then, describing a process that operates cognitively as a way of answering the question: who am I? In this way when engaging and acting on our possibilities, we are no less tied into our future in terms of our
capacity to make choices (Heidegger, 1962:32-33). We live in the present, and
the present is the historical movement of our own past from which the present
is replete with the possibilities of actions that shape one’s own future. An
alternate explanation is to conceive that each one of us (each Dasein) is a
STEM containment-field replete with the intrinsic or essential faculties and
memories, the source embodiment of one’s own actual and interpreted history,
compositionally a temporalising self-organising Being. Primordially, as I
construe, nestled factically as Space-Time-Event-Motion entities, as
constituted energy-motional entities such that each life is an on-going event in
temporal process i.e. Space-Time-Event-Motion (STEM) entities within an
Earth STEM global environment. We do exist in a global time yet we are
constituted embodied autonomic entities having some motional independence
though nonetheless part and parcel of, the planet Earth extending further out on
greater galactic scales. If, as Heidegger maintains existence precedes essence,
then the intrinsic spatio-temporal nature of one’s being requires demonstration.
The spatio-temporal nexus of a being-in-the-world whilst simultaneously being
interconnected to the world of motile Beings requires explanation. My model
of STEM composition and agency fulfils the requirements to arguably resolve
Heidegger’s unanswered questions. The short answer is that humans are,
constitutionally, spatio-temporal beings.

For Heidegger’s account of Dasein’s facticity, refers to its Being-in-the-
world, which makes evident the multiplicity of ways of ‘Being-in’ or engaged
in everyday actualising events (making, breaking, losing, finding, determining,
moving-around, etc.). Heidegger indeed describes three distinguishable ways
that Dasein encounters and realises already interconnected activities: ‘idle
talk’, ‘curiosity’ and ‘ambiguity’. These ‘characterise the way in which, in [an]
everyday manner, Dasein is its ‘there’ – the disclosedness’ of expressive
characteristics these are not present-at-hand in Dasein, but help to make up its
Being’ (Heidegger, 1962:219). All modes of our experience and activities, as
Heidegger construes, are determinant ways of being-in-the-world (Sheehan,
2007:197). Critically observing here contrary to Heidegger, that each Dasein,
phenomenologically speaking, observes their situated environment, reflecting
upon past experiences from a primary and ongoing spatiotemporally oriented
living event as a STEM being in activation so cognised through the myriad
ways of engaging. Determinate ways of being-in-the-world as Heidegger
claims are indeed ways of engagement and activities to which each Dasein
attends but this observation espoused by Heidegger lacks the demonstrative
explanatory power of detailing ‘how’ Dasein temporalises as an already
existent being. Pertinently, in addition each Dasein it must be noted is to itself
a frame of reference constituted, I argue as a STEM agent and is always
already therefore a temporalising embodied observer.

Moving to Heidegger’s perspective regarding the concept of ‘Falling’ as
used primarily to refer to these ways of engaging - ‘idle talk’, ‘curiosity’ and
‘ambiguity’ – explaining ‘there is revealed a basic kind of Being which belongs
to everydayness: ... the “falling” of Dasein’ (Heidegger, 1962:219). Yet ‘Fallenness’, for Heidegger, is not construed negatively but extends the manner in which Dasein may find itself ‘alongside the world of its concern’ (Heidegger, 1962:219). This observation is central to understanding Heidegger’s development of the authentic self because falling takes on the ascription of ‘inauthenticity’ as it were, it:

... has mostly the character of Being-lost in the publicness of the ‘they’. Dasein has... fallen away [abgefallen] from itself as an authentic potentiality for Being its Self, and has fallen into the ‘world’. “Fallenness into the ‘world’ means an absorption in Being-with-one-another, in so far as the latter is guided by idle talk, curiosity, and ambiguity (Heidegger, 1962:220).

Fallenness also indicates a pre-ontological mode of being (i.e. not questioning, living day to day unquestioningly) in which Dasein essentially fails to acknowledge its ontological condition (capacity to question) and instead ‘falls back’ to everyday inauthenticity by falling back into tasks (readiness-to-hand existence) and consequently reverting back into das Man. To explain this notion Heidegger claims, contra Descartes, that the ordinary self is not an individual self, nor the self of Cartesian reflection, instead it is an “anonymous” self one defined by other people. The Das Man self, though essential to life, is not our genuine self but a comparative self assigned by the roles we play and social categories (Solomon, 2000:62).

Heidegger provides an interesting psychological depiction of not being a genuine self though simultaneously resembling a condition of other self-making potentiality such that Heidegger’s analysis uncovers the manner of contrivance as born from social circumstance such that the ‘they-self’ construal once objectified is not a concrete or inflexible self. Essentially Heidegger’s analysis uncovers that what is highlighted from this process of phenomenological examination is the human capacity to exhibit and express different ways or modes of self-being. Witness the adoption and varied manifestation of the social-able self. Yet we are anchored by our spatiotemporal embodiment as the potential ground of authentic being in virtue of our individuated compositions, in many ways cognitively reinforced, through social or worldly community life. Counter-intuitively, it seems that the world of people in Heidegger’s analysis would otherwise be replete with no-self beings only social selves or Dasein’s whilst as the observational examiner nonetheless expects there exists the potential in each individual or Dasein to discover one’s own authentic self though only through the recognition of exiting within a world of its being.

Building on this point Heidegger’s sense of authenticity involves the somewhat foundational notion of ‘care’. Heidegger’s notion of the term ‘care’ (Sorge) is extensively construed to imply being involved, of being ‘concerned’, specific to having ‘ends and purposes’ entwined with Dasein’s potential (Solomon, 2000:62). The term ‘care’ captures ‘being engaged in the world’ and
in our own lives, particularly with respect to our possibilities. We concern ourselves with the things around us and the situations we confront; and in doing so, “we do not just stand in indifferent or inert relations to them” (Blattner, 2006:44). For Dasein then, the concept of care reveals itself in a significant way such that all our activities and who we are, is suffused with caring (Heidegger, 1962:274).

Accordingly when examined from the standpoint of Being-in-the-world Heidegger observes that ‘everydayness reveals itself as a mode of temporality’. In other words our lives always involves time, it pervades every aspect of our lives since everything we do inevitably involves time. Indeed Heidegger subsequently advances a conditional argument: “if temporality makes up the primordial meaning of Dasein’s Being and this entity is one for which, in its Being, this very Being is an issue, then care must use ‘time’ and therefore must reckon with time” (Heidegger, 1962:278). Significantly on closer examination Heidegger equates one sense of time by which “entities within-the-world are encountered-time as “within-time-ness” (1962:274). Through the “uncovering of the inner-time-consciousness of Dasein’s Being-in-the-world” the endeavour to penetrate the ‘inner heart of time’ sees the project move from the Being of time to the time of Being” (Ward, 2008:101). Heidegger contends that the “primary item in Care is the “ahead-of-itself” (living in the present yet simultaneously in a future direction) however explaining that “Dasein exists for the sake of itself” (Heidegger, 1962:279). Reckoning that “[A]s long as it is’ “right to its end, it comports itself towards its potentiality-for-Being”. The ‘ahead-of-itself’ as an item in the structure of Care discloses that in Dasein something is left outstanding, “not yet become actual” (1962:279).

Heidegger’s construal of this analysis is in terms of Dasein achieving ‘wholeness’ though in doing so he incorrectly inferred would otherwise liquidate its Being-in-the-world as capable of being experienced as an entity (1962:280). Ontologically the quest consequently became characterising Dasein’s being-at-an-end and of achieving an “existential conception of death as an existential projection of an authentic Being-towards-death” (1962:281). Though this observation is very significant and not to be undermined it however ushers in a phenomenological shift. In summary my contention is that to be a ‘being-there’ (Dasein) as a Being-in-the-world is to be an actualising temporal being hence a STEM embodiment first and foremost. Even Heidegger did not believe that time was a thing in and of itself independent of the world and specifically not independent of Dasein. Moreover Heidegger claims that “[B]ecause Dasein as temporality is ecstatic-horizontal in its Being, it can take along with it a space for which it has made room, and it can do so factically and constantly” (Heidegger, 1962: 420ff). This is equivalent to saying that the ecstasies inherent in Being – living through past, present and future temporalities involves the space we inhabit. In Heidegger’s analysis where Dasein takes in space relates to the engagement with equipment relative and limited to a fallen state; such to bring something closer (e.g. in its use) one moves the thing from its “thence” and as such “making-present forgets the
yonder”. But this analysis also demonstrates what is crucially lacking in Heidegger’s overall analysis and that is the perspective of embodied space. Instead Heidegger seems transfixed on the ecstatico-horizontal temporality as developed in those following pages of Being and Time.

Heidegger endeavours to analyse the coupling of space and time but never really captures the conceptual, let alone the ontological understanding, that being can be thought of as the spatialisation of time (spacetime) and together the enactment of the process of existence within the world/universe. Space and time are interconnected as Einstein’s theory of relativity has resoundingly demonstrated. One need only look out into the night sky at the distant stars to realise one is also looking back in time.

Following a different path however Heidegger in Being and Time, claimed that Dasein’s existence, facticity and falling reveal themselves in the phenomenon of death (or Being-towards-the-end) (1962:293). Falling, in this sense as Heidegger describes, is a kind of fleeing in the face of death such that being-towards-the-end has a ‘mode of evasion’ in the face of it, significantly for this examination, which takes the ascription of inauthenticity. Dasein, we recall is constituted by ‘disclosedness’, an understanding equated with a state-of-mind such that to ‘comport’ oneself towards death is “Being towards a possibility of Dasein itself, as the possibility of authentic existence” (1962:307). It follows that the manner in which one comports oneself is central to achieving authenticity. Not surprising that the characterisation of being-towards-death involves anxiety (a state-of-mind) and ordinarily anxiety is directed towards cowardness. Interestingly, however, Heidegger sees this construal of anxiety as a perverted state (1962:311). Dasein’s selfhood is taken formally as a ‘way of existing’ and not as an entity present-at-hand. Heidegger claims that the ‘I’ for most part is not the ‘who’ of Dasein, he says instead “the ‘they-self’ (or Das Man) is its who” (1962:312). Heidegger contends that “Authentic Being one’s Self takes the definite form of an existentiell modification of the ‘they’ (that is, as defined by what one does or whatever designation, baker, waiter, candlestick maker, etc.) announcing that this modification must be defined existentially” (1962:312). However, it seems that this way of understanding the self makes it a relative term, perhaps, as intended. For when Dasein brings itself back from the “they” (i.e. the socially assigned self), the “they-self” is modified ... it becomes authentic Being-one’s-Self. Apparently, since Dasein is lost in the “they” (simply conforming to the social self in an unquestioningly manner), it must first find itself and it can find itself only because it has already in itself, possible authenticity (1962:313). Most significantly what still remain undiscovered through this analysis are the substantive (factual) constitutive elements against which the relative authentic Dasein rests and not arguably the adapted version of the juxtaposed Das man self or ‘they self’.

As we see for Heidegger the “I” is rather the subject of logical behaviour of binding together, such that “I think”, means I bind together. “All binding together is an ‘I bind’ together” (Heidegger, 162:367). I bind together from the factual state as I contend enabled only as a constituted STEM. Heidegger
believes that the *subjectum* is ‘consciousness in itself’, not a representation but rather the ‘form’ of representation. The ‘I think’ says Heidegger, “is not something represented, but the formal structure of representing as such, and this formal structure alone makes it possible for anything to have been represented” (1962:367). Essentially Heidegger is contending that the ‘I’ as articulated as the *res cogitans* is not pointing to a substance as present-at-hand but instead to a process as “I think something” (1962:367). *Dasein* becomes essentially *Dasein* in so far as authentic existence constitutes itself as anticipatory resoluteness; and as such resoluteness, as a mode of authenticity of care, is thought to contain *Dasein*’s primordial Self-constancy and totality (1962:370). This of course contrary to Heidegger’s insistence is an attitudinal shift in terms of self-reflection and self-examination. However, if this primordial nature is not construed in any substantial manner what is not clear in this analysis is whether there is a distinction between mind as ‘I’, and mind as *Dasein*, though not an ‘I’ substantively construed but nonetheless recognisable simply, as Heidegger insists, a process?

In one sense we are told it is ‘consciousness in itself’ as the ‘form of representation’. Yet in another sense somewhat conflated what is recognised is a process; but of what, consciousness? Surely one does not need to completely disengage from the world to be able to disengage from the constructive process of the ‘I’, of thought. Heidegger says ‘I binds’ and as such what he identifies is the process of it grasping, seizing, in making meaning by constantly attaching to objects in the world in an identifiable manner of which he himself distinguishes two categories of Being (readiness-to-hand and presence-at-hand). Detaching from the mind that is to say not focusing on the content of one’s thought as when in deep state meditation for example, is however not a detachment from existence. Instead it can be a full emersion of authentic being, an understanding that has a long history in many Eastern meditative traditions. Given that Heidegger’s analysis of the conscience is consistent with the idea of nullity (no-thing-ness) that nonetheless in-forms as the ‘call of conscience’, arguably is equiprimordially of what I have previously argued is one’s unique Signature-Energy-Frequency (SEF) (Principle of Individuation) (Naimo, 2011). An analysis quite consistent with Heidegger’s analysis of *Dasein*’s Being-in-the-world but in this case expanding the conceptual description of *Dasein* in terms of embodied STEM agency.

Heidegger suggested that temporality reveals itself as the meaning of “authentic Care” (Heidegger, 1962:376). Time goes on, he describes, and the authentic future is ‘the towards-oneself’. “*Dasein* is historical” (1962:381) and for Heidegger, this construal is significant in that its ascription identifies an aspect of being which differs from simply being part of history, as a part of some recorded event in history. Heidegger employs the term ‘Historicity’, not construed as history, but instead as the condition of being historical, of *Dasein*’s own life-history in an embodied manner. Connected to this is the understanding that *Dasein* is a being-in-time for “*Dasein* uses itself up”; that is, uses its time and in using time *Dasein* reckons with it. Reckoning with time is
Temporal Being and the Authentic Self

constitutive for Being-in-the-world Heidegger observes (1962:382). The notion of existence against this construal is conceived in the sense in “which one’s Dasein exists as a coming-towards-onself such that projection is basically futural” (1962:385). In one sense projection is the structure of understanding and so to understand something is to project it. This point of futural projection can only make sense if time is constitutive of being, more integrally, as I argue, in STEM composition. For what else could be the reference point for Dasein? Since for most part Heidegger construes the sense of authentic present against a counterpart inauthentic present. “That Present, which is held in authentic temporality and thus is authentic itself”, Heidegger calls the ‘moment of vision’ – understood in the active sense as ecstasies (rapture of three temporal dimensions co-existing) (1962:387). Augenblick in Being and Time is translated as ‘moment of vision’ taking the ascription “related to ‘sight’ or ‘insight’ [Einblick]. ...the Augenblick is the ‘vision’ into Being which reveals Dasein in its ownmost possibilities of Being, and through which Dasein can experience an extraordinary and ‘totalizing’ sense of Being” (Ward, 2008:112).

States-of-Mind

“Understanding is never free-floating”, but is associated with some state-of-mind. The ‘there’ gets equiprimordially (is foundational) disclosed by one’s mood in every case or gets closed off by it in falling (Heidegger, 1962:389). Having a mood brings Dasein with its thrownness disclosed far more primordially in ‘how one is’”. “Existentially “Being-thrown” means finding oneself in some state-of-mind or other” and Heidegger concludes that one’s “state-of-mind is therefore based upon thrownness” (1962:389) somehow just from being there as our first realisation. One’s mood takes on the ascription of representation in realising ‘whatever may be the way in which’ one is ‘primarily the entity that has been thrown’ (1962:390). Heidegger explains: Bringing Dasein face to face with the “that-it-is” of its own thrownness – whether authentically revealing it or inauthentically covering it up – becomes existentially possible only if Dasein’s Being, by its very meaning, constantly is as having been. ... the ecstasis of the “been” is what first makes it possible to find oneself in the way of having a state-of-mind”. The Understanding is grounded primarily in the future; one’s state-of-mind however, temporalises itself primarily in having been. Moods temporalise themselves – that is, their specific ecstasis belongs to a future and a Present in such a way, indeed, that these equiprimordial ecstases are modified by having been” (1962:390). Being-in-the-world, in the most inseparable way, is rooted in the spatiotemporal domain of existence. Moods then are regarded as “fleeting experiences” which ‘colour’ one’s whole ‘psychical condition’ and because of this fleeting character Heidegger claims “belongs to the primordial constancy of existence” (Heidegger, 1962:390). Past experiences in other words act as the ground upon which the present, in living towards-the-future in act is the checkpoint for choices made. The construal that one’s state-of-mind is
grounded primarily in ‘having been’ means that the “existential basic character of moods lies in the bringing one back to something” as if to say something central about one’s being, of one’s core (1962:390). Genetically speaking each individual is relatable through their own unique genetic makeup. Likewise one level of analysis is the construal of an entity identifiable in terms of one’s own unique signature-energy-frequency (SEF) (Naimo, 2009:191).

Answering Heidegger’s Unanswered Questions

*Being and Time* we know was an incomplete work to which Heidegger in the final section of the work returned to the question of the meaning of being. At the very end he raises these questions after recognising that:

“The existential-ontological constitution of Dasein’s being is grounded in temporality. Hence the ecstatical projection of Being must be made possible by some primordial way in which ecstatical temporality temporalizes. 1) How is this mode of the temporalizing of temporality to be Interpreted? 2) Is there a way which leads from primordial time to the meaning of Being? 3) Does time itself manifest itself as the horizon of Being?” (1962:488).

For the first question: How is this mode of the temporalizing of temporality to be Interpreted? To answer this question I argued throughout requires a perspective shift such that space and time are regarded as mutually interdependent as in the model provided in this paper i.e. Space-Time-Event-Motion (STEM) containment-field embodiment (human being); and previous works where I have laid the foundation for this conceptualisation. The second question: Is there a way which leads from primordial time to the meaning of Being? At this stage of human development this is the most challenging question. One can however advance by inference to best explanation what is conceivably an appropriate hypothesis in the manner of a principle of individuation and identity: Signature-Energy-Frequency (SEF) principle. The evidence already exists supporting the plausibility of this account. Frequency, in the sense of resonance is fundamental in Nature associated with all forms of communications: auditory, chemical and relative energy expression.

To elaborate, matter according to relativity theory is inter-convertible to energy as was made famous by Einstein’s mass-energy equation $E = mc^2$. The fabric of the Universe according to this observation is physical and as such is reducible to energy. I draw on the observation of Planck’s constant which is a formula used by physicists to describe the constant proportionality between the energy emitted or absorbed by an atom and the frequency of emitted or absorbed light (Jibu & Yasu, 1995). Energy characterised as such is the fundamental substance of the Universe definable in terms of frequency. Here, then, is one way to think about how primordial time could lead to an
expression and to the meaning of Being as Heidegger hoped to answer. Humans are born within the world from pre-existing beings, i.e. our parents. We, as all things have evolved! Common to all materialised things is an inherent signature-energy-frequency (SEF). Developmentally through our animated conscious engagement with the external world of phenomena our observations are largely turned outward via our sensory channels whilst concurrently much of how we perceive the world happens internally. That is, as spatiotemporally embodied entities, invariably made of the same stuff that exists in the external world we inhabit, we are nonetheless intrinsically part of, in a compositional way, sensing out from within. There is a coexisting mutually interdependent external spacetime with a coexisting interdependent internal spacetime. For Being involves the inner interpretive and organisational aspects conditional upon spatiotemporal sensory modes. As animated spatiotemporal embodied beings the external world of sensory information constantly requires being transformed into sense receptive, ideational or conceptual meanings. What cannot be answered here is how the fundamental energy of the universe transforms itself into the manifold entities and elements that make up the universe.

For the third question: Does time itself manifest itself as the horizon of Being? I answer yes. Explaining in sum by saying ‘Being’ and ‘Time’ is the union and process of existence. Being is the spatialisation of time mutually interdependent, the relation of the enacting process of embodied existence in one’s STEM being. The horizon of projected temporality is consistent with being a STEM constituted being. Heidegger, as the title of his work exemplifies examined ‘Being’ and ‘Time’ as added, somewhat to which we exist in time, just like occupying space but only for a period of time. ‘Being-in-the-world’ was so significantly part of what it is for Dasein to exist that Heidegger, I believe, missed the crucial point of connecting ‘being’ ‘with’ ‘time’ so that ‘Being’ and ‘Time’ are coexistent as they are of course.

On the subject of Authenticity as developed in Being and Time refers to a “way of relating to our existence” described with no specific content, no universal platform issuing from every authentic Dasein (Carman, 2007:289). Heidegger shifts the idea of human existence understood as a unity with a phenomenological account of “Dasein owning up wholly – that is wholeheartedly – to itself in its existence” (Carman, 2007:289). Hence on this account to own up to oneself in one’s existence is to exist authentically such that the deep structure of human is revealed as falling thrown projection (Carman, 2007:289). As such then being authentic is to stand resolute against the Das man (the ‘One’), of the general everydayness. In authenticity the “public understanding of my world” is used by “projecting on my own possibilities” (Dreyfus & Wrathall, 2007:8) achievable it is contended as STEM causal agents. The role of conscience calls to turn Dasein into the reticence of its existent potentiality-for-Being characterised as a resonant sense of understanding equiprimordially, I argued, of one’s Signature-Energy-Frequency. Finally Heidegger’s account of authenticity remains somewhat undeveloped though providing an excellent basis to work from if
reconceptualised in the manner undertaken in this paper. The task now then is to build on this framework.

References


In this paper I argue that Aristotle misconceives politics in that he adopts a hierarchical set up which governs his virtue ethics within a state. The social hierarchies exist as unquestioned premises to regard correctness to an already right conduct by appealing to his system of virtues. I shall substantiate this by explaining the problems associated with the hierarchy within Aristotle’s conceptualization of praxis versus poesis, and phronesis versus Sophia with regard to the connection between ethics and politics. In answer to the question that what type of explanations the domain of ethics is susceptible to I argue that political and moral theories need not appeal to unchallenged premises. In this sense Aristotle’s understanding of politics is not a sufficient one. He presents a narrow view of politics. Political actions on the other hand are free and not determinable by any ends.

Keywords: Politics, Virtues, Praxis, Eudaimonia, Sophia, Phronesis

Introduction

The knowledge of ‘good’ and its practical application in life are the two parts of ethics according to Aristotle. For him philosophical knowledge of ‘good’ is a ‘kind of political science’ (Nichomachean Ethics, 10) and it is the responsibility of state to ensure the practice of virtues by its citizens. In ‘Politics’ he provides various categories of ‘good’ of each person depending on what one is by stature in social hierarchy like the good of a woman or that of a slave or a ruler which are all different from one another depending on role they play in the social structure. Most philosophers have disagreed from this view of Aristotle connecting ethics and politics wherein state is run by a pre decided hierarchy of virtues based on a role for everyone within the social matrix. This is problematic because given an unjust hierarchy even the most ideal state would fail to be moral.

Common Experience about Good and Beautiful are Treated as Facts and as First Principles

In the first chapter of The Nichomachean Ethics, Aristotle describes happiness as ‘living good or doing Good’ (7) or ‘the virtuous activity of the
soul’ (15). This practice of good life is based on first principles or commonly known ‘facts’ and in case of Aristotle’s The Nichomachean Ethics, where he discusses virtues his first principles are the commonly known facts of his time about qualities that are considered virtuous and those that are counted as vices. He never questions them and the same applies to vices as well e.g. he takes for granted that adultery is a vice or courage is a virtue. In the first chapter of Nichomachean Ethics, he writes:

‘We must start from what is known. But things are known in two senses: known to us and known absolutely. Presumably we must start from what is known to us. So if anyone wants to make a serious study of ethics, or of political science generally, he must have been well trained in his habits. For the starting point is the fact: and if this is sufficiently clear there will be no need to ascertain the reason why. Such a person can easily grasp the first principles if he is not in possession of them already.’ (8).

Later on he quotes Hesiod to make it clear that learning of ethics is based on known facts about ‘good’ and ‘beautiful’, either acquired or known to us. There is no need to argue if things are known as common experiences but this leads to cultural relativism since value recognized in one time zone may not be true across all times and all places. His approach to ethics reduces it to mere conditioning of socially accepted virtuous ‘habits’ which he never tries to question but accepts it as a known ‘fact’ since in Greek times it was known to everyone who was virtuous and what constituted of virtue. But this does not hold consistent in the present times. Also since these known facts about ‘good’ are not accessible to many people like those who remain uneducated so there are no possibilities of being ‘good’ or moral for them. He excludes a large section of the human population from having a ‘good’ or ‘moral’ life except for the elite few who are trained in the facts about ‘good’ and who can practice them. Aristotle implies here a social hierarchy in his approach.

**Politics is the Life of Action for the Sake of Necessities**

He further connects the practice of a virtuous life to three aspects of human life namely pleasure, action and contemplation. He regards pleasure and action to be inferior to the life of contemplation because they can be performed even by animals or gods but it is thought which makes human beings different from animals and gods. Now here Aristotle makes an underlying distinction between labor of human body, work of limbs and the activity of human mind which he calls contemplation. He separates labor, work and thinking in watertight compartments without allowing any exchange from each other. For example there are some actions which involve ‘contemplation’, not just by the solitary individual (as in case of the creativity in the goal oriented work of a
craftsman) but which are facilitated by exchange of thoughts of others through free speech and have no set goal but they are exceptional in that they are stepping stones to arriving at the unachieved by human beings like in case of scientific discoveries that are triggered by various such contributions. They cannot be judged by their consequences but by mere potential/courage of the doer for such exceptional actions. But Aristotle here does not give due importance to the fact that political actions are not acts of necessitation as the acts of work like those of craftsmen, painters, dancers etc. which can be judged by the aptness of their outcome. They are not purposive or creative in the sense of bringing out a finished product like writing a story or conducting elections or administration but they are unpredictable and free. A consequence following from such categorization by Aristotle is that the life of philosophers becomes superior to the inferior kind of life of those involved in political actions while the rest do not have a lived life at all.1

**Praxis versus Poesis**

Within Greek philosophy there is a distinction between *praxis* or doing and *poesis* or making. Praxis is different from production so all the acts of workmanship or creating artifacts, skills are not praxis. Aristotle describes human life as that of ‘praxis’ that is self-referential and in order to reach disclosure of ‘who’ someone is it is not sufficient to know what someone has made in a life. In order to have this disclosure such individual actions must be witnessed by others and we must tell and retell stories about the ‘who’ in question from various perspectives or we may say praxis survives in stories as told by other people. The ‘who’ remains hidden from the person ‘like the *daimon* in Greek religion which accompanies each man throughout his life, always looking over his shoulders from behind and thus visible only to those he encounters’ (179-80) It is for the preservation of this praxis, polis was created as a protected space of good world/life (unlike the household which is the domain of necessity) where praxis can survive and can be performed and thereby overcome the transience of human life.

Aristotle assigns an end to *praxis* also as ‘itself’. It is akin to a ‘skill or *techne*’ whose end is ‘itself’ as opposed to the ends of production or making in case of *poesis*. Aristotle regarded political actions as acts of ‘making’ this is why he differentiated between *praxis* and *poesis*.

While to attribute an end to political actions would be to undermine the fact that apart from the individual dimension of the doer of these actions there is also a social dimension of political actions constituting of the others who witness these actions and respond to them. This social dimension of political actions can never be controlled or pre-decided by any doer which makes political actions stand outside any individual’s control. Political actions are free from the controls of any ends in this sense. This is one of the key distinctions between actions of workmanship and acts in the political domain.

---
Conflict between Philosophy and Politics (vita contemplativa versus vita activa)

The Greek city states erroneously ‘essentialized’ political actions so that it was only thinking that could free human beings from the burdens of essential demands, both of body and society. But before that the meaning of Polis was not of an organization that mandated political actions rather it was one which facilitated political actions to ensure freedom. To enforce the performance of political actions as compulsion is against the very meaning of political actions which ought to be free in their origin. Political actions can never be foretold before their performance. This was an error which was committed by Plato and Aristotle who imposed this hierarchy between thinking and political actions.

In this sense there is a conflict between philosophy and politics the same way as the conflict between eternity and immortality. By performing Political actions human beings strive for immortality in their exceptionality expressed either in language, speech or action, while eternity is the domain of thinking, an experience which cannot be described because description will take it into another realm which is contrary to it (Socratic school believed in this). Eternity is a domain of theoria. While immortality can be achieved by humans through remembrance of exceptional deeds performed in one’s life by future generations but eternity remains an unachievable realm of the ‘divine’. In the cave parable Plato proved that philosophers’ experience is akin to death and is eternal. He argued that the world of political actions is not enough but a higher principle was possible which stands above the formative principles of polis and this was the realm of theoria or contemplation where eternal experience was possible and every human act was subordinate to it. Contemplation is directed to itself and has no other purpose but likewise politics also does not have any purpose since it is not for achieving any goals but for itself. However Aristotle describes politics as directed to achieve ‘public honor’. But this distinction is based on a false conception of Polis as essential while otherwise there is no hierarchy between politics and philosophy (theoria and praxis). It was erroneously imposed by Plato and Aristotle like Socrates too who chose a life of eternity and contemplation by refusing to write his books which would have ruined the experience of contemplation.

Virtues and the Problem of Flourishing

Aristotle gives a catalogue of virtues as a list of good behaviors without organizing them in a meaningful way of any justification as to why they should or should not be followed. Aristotle makes the fallacy of begging the question by arguing that virtues are needed for living well (he has already assumed what it is to live well as a given and not a conclusion) and not the other way round i.e. that life is lived well through virtues by giving any justification for the quality of life through virtues. He argues that virtues help in acquiring ‘eudaimonia’ or the greatest ‘good’ often wrongly translated as ‘happiness’ but
it means ‘flourishing’ or ‘living well’ according to one’s potentialities using excellences which act as tools to achieve the purpose of flourishing, e.g. a mango seed has the potential to grow into a mango tree and it has some excellences (organs in the plant itself) that can help it to flourish into one so arête exists as tools to aid ergon. But neither does he describe what constitutes ‘flourishing’ (it is only a formal description) nor does he detach ‘eudaimonia’ from its metaphysical connection when he says that (it is ‘precious and perfect’ and the first cause of it is ‘divine’) it is unachievable in life ‘because we call the Gods ‘happy’ or ‘blessed’, a term which we apply also to such men as most closely resemble the gods.’ (26). Besides Aristotle also insists on the social determinant of ‘living well’ or ‘flourishing’: ‘Happiness [here it is a translation of eudaimonia] is such an end which is chosen for its own sake. This self sufficiency of happiness is ‘not for oneself alone living a solitary life, but something that includes parents, wife, and children, friends and fellow-citizens in general; for man is by nature a social being’ (13) For him knowledge of virtues is a pre-condition for living well. As he also writes that children and animals cannot be happy which suggests that he refers to only the regulation of socially accepted values/virtues by a process of consistently perfecting them in practice throughout the life of an individual only then one can create the possibilities of living well or ‘eudaimonia’ which continues to be unachievable even after going through such a perfection in life.¹

**Eudaimonia is Unachievable**

The concept of ‘virtue’ is further based on two assumptions 1) that of teleology i.e. human life has a function/purpose which he calls ergon (a noun with multiple connotations of function, task, work, tool, excellence) and 2) It is socially determined.

Aristotle writes, ‘the function of man is a kind of life, namely, an activity or series of actions of the soul implying a rational principle’ (16). So the function of man if there is one such function, it is to act in accordance with the rational principle (the rational part of the soul) using one’s excellences and it is socially determined.

In *Politics*, Aristotle describes that ergon varies according to one’s arete (excellence) and if there are numerous arete (if one excels in many like a horse can run and also gallop) then according to the best and most complete. He explains that arete (Agathos) is the excellence or good of human being and such excellence is not of the body but of the soul. He writes: ‘The goodness that we have to consider is human goodness, obviously; for it was the good for man or happiness for man that we set out to discover. But human goodness is meant goodness not of the body but of the soul, and happiness also we define as an activity of the soul.’ (28). Some knowledge of the soul is needed in order to understand one’s arête however the concept soul in Aristotle’s writings

¹Martha Nussbaum in ‘Fragility of Goodness’, 375-377 argues that Aristotle’s concept of Flourishing is not a unified concept and that there are contradictions within Aristotle’s account of as to what constitutes Eudaimonia.
pertains to a metaphysical biology which he explains in De Anima and his writings on Metaphysics.\(^1\)

Since goodness is a metaphysical concept so it cannot be achieved by human beings but only by god. While the philosopher can live in company of thoughts in solitude but even a philosopher cannot be alone like the God who is alone in goodness because good deeds must be forgotten the moment they are performed so there is no body to accompany the god. Moreover goodness is always in the realm of the private or secret because when it is exposed in public it no longer remains good as its character changes with public expression of it (depending upon its usage) so it is anti-public. So goodness cannot be a part of politics.

**Prudence and the Virtues**

His idea of ‘virtue’ is that of ‘a purposive disposition, lying in a mean that is relative to us and determined by a rational principle, and by that which a prudent man would use to determine it’ (42). It is prudence (*Phronesis*) that helps in deciding the mean i.e. what is extreme or deficit in regard to a particular virtue in context. So both rationality or intellect and prudence (practical wisdom) are at work here. The difference between rationality and prudence is that while rationality is based in reason, prudence is a kind of ‘perception’ which is not based on reason: ‘prudence ‘apprehends the ultimate particular, which cannot be apprehended by scientific knowledge, but only by perception-not that of objects peculiar to one’s sense but the sort by which we perceive that the ultimate figure in mathematics is a triangle’ (156). It directly perceives virtue rather than calculating it by reason. But Aristotle makes clear that prudence is necessary to control wisdom: All virtues are qualified ‘in accordance with the right principle’; and the right principle is that which accords with prudence’ (165). It is the virtue of the calculative part of the soul. It is not possible to be prudent without the possession of all other virtues. It is prudence that controls all wisdom. Aristotle further explains that all virtues not only conform to prudence but imply it. But if prudence is a ‘perception’ it cannot ‘subsume’ any general principles into a particular but it can only perceive correctly. The possession of all virtues is a precondition to have prudence so it cannot be possessed by everyone so it a very circular process to be virtuous in Aristotelian sense since to know the mean one needs prudence and in order to have prudence one must possess all virtues. So those who do not know of the facts about ‘good’ or are not already virtuous will never live well. Aristotle’s approach is very exclusive in this sense.

**The Possibility of Going Wrong in Calculating Virtuous State**

Further the concept of virtue is stated only in outline as a formal concept with no specific content which Aristotle writes is relative to context and also varies from time to time and it is always in relation to a particular person. He

\(^1\)A.W.H. Adkins explains these concerns in his paper ‘The connection between Aristotle’s Ethics and Politics’
writes: The matters of conduct, ‘must be stated in outline and not in precise
detail’ because it varies from one situation to another so we can only say about
behavior in general terms: ‘the agents are compelled at every step to think out
for themselves what the circumstances demand, just as happens in the art of
medicine and navigation’ (33). One critique that follows from it is that if a
person rationally calculates injustice or cowardice to be the mean then within
Aristotle’s domain one is not able to resist it since there is no content to his
concepts of happiness and virtue as he leaves it for every person to decide it on
one’s own through rational calculation in reference to one’s character.

**Types of Virtues and the Golden Mean**

Within Aristotle’s analysis virtues mark a reference to potentiality and
actuality: ‘of all those faculties with which nature endows us we first acquire
the potentialities, and only later effect their actualization’ (31). Given the
potentiality for a virtue, we acquire it by first exercising it. The behavior and
actions make one of a particular disposition: ‘like activities produce like
dispositions’ (31). Unlike the virtues of the body which nature has endowed in
us, Aristotle here refers to the virtues of the soul which cannot be acquired
unless they are put in practice even if one has the potential for them e.g.
intelligence or courage does not contribute to one’s development unless it is
performed through actions, even though to have these excellences still counts
virtuous for Aristotle. Virtues are like tools and by acquiring these tools human
beings take on to the *ergon* or purpose in life which is to flourish in accordance
with their soul.

Virtues are of two types:

1) Intellectual (wisdom, understanding and prudence)
2) Moral (Courage, temperance, liberality, magnificence,
magnanimitiy, proper ambition, patience, truthfulness, wittiness,
friendliness, modesty, and righteous indignation).

Intellectual virtues refer to a virtuous state of mind whereas moral virtues
describe a virtuous character. However Aristotle’s classification between moral
and intellectual virtues does not hold for virtues like wisdom which are both
moral and intellectual. Besides, without acquiring prudence one cannot become
virtuous while prudence requires the possession of all virtues which has
already been discussed as a circular argument used by him.

**Phronesis and Sophia**

Aristotle puts contemplative life as superior to the life of action or politics.
However, politics is governed by *phronesis* contemplation is ruled by *theoria*
or wisdom (Sophia). While *Sophia* relates to the divine, it is *phronesis* or
practical wisdom which concerns human beings. *Phronesis* constitutes *praxis*
in agreement with *wisdom* (well thought out action), it is developed by
plurality of perspectives in the realm of human actions unlike *Sophia* which is theoretical in nature.

A prudent man is one who can calculate ‘what is good and advantageous’ for him not in matters of any particular conduct but ‘to a serious end’ that is, ‘what is conducive to good life in general’ (150). Prudence cannot be science since it not possible to deliberate about necessary things that is the object of science. It is not art which concerns production as prudence relates to action or ‘doing well’. Prudence is a virtue that concerns the calculative part (one that forms opinions) of the rational soul. It is ‘a true state, reasoned and capable of action in the sphere of human goods’ (151) similar to temperance (self-control). There is no excellence in case of prudence whereas a man who mistakes in prudent action is rated lower.

On the other hand, to be wise or to possess *Sophia* one must possess both the knowledge of the first principles and ‘a true understanding of those principles. Therefore wisdom must be intuition and scientific knowledge: knowledge ‘complete with head’ (as it were) of the most precious truths’ (153). Wisdom is different from prudence (or what is beneficial for being) since there cannot be any one common good that applies to all beings on earth. What is prudent for a fish is distinct for what is beneficial for a human being. Unlike prudence which aims at ‘practical good’, the objects of search for the wise are not ‘objects of human good’ but it is precious knowledge that does not aim at usage (useless but of high importance). Prudence concerns both the knowledge of both universals and particulars in order to administer the conduct of practical affairs.

‘Politics is the fullest realization of prudence’ but ‘their essence is not the same’. Prudence concerning the state has two aspects: legislative science and political science. Legislative science relates to controlling and directive while political science which is practical and deliberative, relates to enactment. Besides, prudence is also identified with self and individual apart from all other forms like domestic, legislative, political science, juridical science. A prudent person seeks his/her own ‘good’ and confines to one’s own interests. But for Aristotle ‘it is impossible to secure one’s own good independently of domestic and political science’ (155). However, it is possible to achieve excellence in understanding universals at a young age but prudence also includes the knowledge of particulars which comes through experience.

Prudence is opposite to intuition. While intuition ‘apprehends the definitions which cannot be logically demonstrated’ prudence ‘apprehends the ultimate particular, which cannot be apprehended by scientific knowledge, but only by perception—not that of objects peculiar to one’s sense but the sort by which we perceive that the ultimate figure in mathematics is a triangle’ (156). This perception in case of intuition is not sense perception but intellectual.

Aristotle creates a rift and a hierarchy between *phronesis* and *Sophia* by attributing different domains to them. The domain of *Phronesis* is politics which is goal directed for Aristotle. However I have already argued that to reduce politics to administration and management is to negate politics itself
thereby Aristotle’s hierarchy between Sophia and Phronesis on these grounds do not hold very convincing.

Conclusion

In his explanation Aristotle creates a flawed hierarchy between praxis and poesis; between phronesis and Sophia which remain unchallenged within his writings and obscures current understanding of these concepts. This hierarchy between politics and philosophy reinforces the understanding of politics as an inferior and subordinate enterprise in the modern society however the roots of this conflict lie in the form of nation building which recognizes politics as essential and pre-decided set of actions/duties for administration and house-keeping of the State and puts the erroneous arrogance of ‘contemplation’ in its stature as something higher. Nevertheless both nation building and the superiority of contemplation over acting are false. They are exclusive to any form of politics.

References

Assuming a Christian perspective, Kierkegaard argues that our moral lives aim at perfection and that we must of necessity fail to achieve that perfection. This consequent failure in ethical life makes possible the existential choice of faith in God as the Person who can forgive such ethical failure.

Against Kierkegaard’s usage of the quotation from the Gospel according to Matthew: ‘Be perfect, therefore, as your heavenly Father is perfect (NIV, Matt 5:48),’ this paper follows scripture scholars who have found that ‘Be perfect’ means not ‘Be morally flawless, but to love as God does, maturely and compassionately, without partiality,’ a command that is possible to realize at times. In contrast with Kierkegaard’s assumption of an ethics that demands perfection, this paper takes up Kierkegaard’s own suggestion that an Aristotelian ethics need not aim at an impossible perfection in its definition of virtue. Furthermore, the paper finds a deep emphasis upon compassion in Aristotle’s analysis of friendship of the virtuous, suggesting that forgiveness of a friend for a moral flaw is possible for the sake of helping a friend to become virtuous again.

The paper then considers more fully Kierkegaard’s arguments (1) that everyone fails the ethical and (2) that the ethical sphere cannot forgive significant ethical failure. In final response, the paper argues that it is reasonable to conclude that the ethical level can incorporate loving forgiveness as a basic attitude towards self and others which brings about personal growth.

We will take up the argument of Kierkegaard that appears to be rooted in a Christian ethics that emphasizes how humans always fail to live up to the ethical command. Kierkegaard identifies the ethical stage of life as an absolute requirement for choosing the religious stage of life in that a person must have passed through the ethical stage and failed at it so profoundly that one needs to choose to believe in God as the only way of forgiving one’s sin (Kierkegaard, 1941, 347 and 26). This point might be easy for some persons to affirm for themselves since they may believe that they have committed an unforgiveable sin and that only God can forgive them. However, Kierkegaard’s argument is intended to be universally applicable to every human. He argues that every
person must seek the ethical and necessarily fail in doing so. As he writes: ‘The ethical begins straightaway with this requirement to every person: you shall be perfect; if you are not, it is immediately charged to you as guilt (Kierkegaard, 1967, 998).’ The assumption of Kierkegaard is that moral perfection is moral behavior which is flawless in the execution of the deed and in the purity of the moral intention behind the deed.

From a Christian perspective, Kierkegaard assumes that ethics deals with the attainment of perfection. The quotation on perfection is from Jesus in Matthew 5:48, as the New International Version states: ‘Be perfect, therefore, as your heavenly Father is perfect (NIV, Matt 5:48).’ However, this Kierkegaardian interpretation, although it fits with what many others interpret, does not fit with all Biblical scholarship. Fred B. Craddock, professor of preaching and New Testament at Candler School of Theology in Atlanta, comments:

It helps to attend more carefully to the word "perfect." The word does not mean morally flawless but rather mature, complete, full grown, not partial. Luke uses the word to speak of fruit maturing (8:14) and a course being finished (13:32). John uses it to describe the fully realized unity of Jesus’ followers (17:23) and James employs the same word to characterize works as the completion of faith (2:22). Paul’s favorite use of the word is to portray the quality of maturity among Christians (I Cor. 2:6; Eph. 4:13; Phil. 3:12, 15). However, this command to be perfect comes most clearly into focus and into the realm of reasonable expectation when viewed within its context. First, the call to perfection comes within a discussion of relationships. Second, Jesus rejects for his followers relationships that are based on the double standard of love for the neighbor and hatred for the enemy. . . .

. . . God does not react, but acts out of love toward the just and unjust, the good and the evil. God is thus portrayed as perfect in relationships, that is, complete: not partial but impartial. God’s perfection in this context is, therefore, love offered without partiality.

Jesus calls on his followers to be children of God in this same quality. "You, therefore, must be perfect, as your heavenly Father is perfect." In other words, you must love without partiality, as God does. Thus understood, perfection is not only possible but actually realized whenever and wherever our relationships come under the reign of God [that is, under the reign of the supreme moral agent who loves without partiality] (Craddock, 123).

Furthermore, there is another version of the Sermon on the Mount from Matthew 5: 48, which appears in Luke’s Sermon on the Plain, Luke 6:36, and is closer to the original Aramaic of Jesus: ‘Be compassionate as your Father is compassionate (CEB, Luke 6:36).’ Richard Gula points out that ‘Compassion comes from the [Aramaic] word for “womb” (Gula, 103) He also notes that
‘Elizabeth Johnson further illuminates divine compassion to be like a mother moved by concern for the children of her womb (Gula, 103).’ This translation of Luke certainly brings out the all-encompassing, impartial love which God has for all, both the just and the unjust, and which is to be the model for human behavior and fits very well with the insightful commentary of Craddock on Matthew’s usage of the word ‘perfection.’ The point is not the absurd demand that humans be perfected in every moral detail of their lives, but that humans be compassionate as their Father is, a command that is possible to realize at times.

However, when Kierkegaard argues for a strict notion of perfection as something which humans are called to do, but which it is impossible for them to do, he is also assuming a harsh and judgmental ethics. Ethics is like the law, a strict disciplinarian that demands absolute obedience to the difficult command of perfection, but which is an impossibility (Kierkegaard, 1980, 16-17). However, although Kierkegaard does not prefer an Aristotelian ethics, he himself points out that Aristotelian ethics does not aim at an impossible perfection in its definition of virtue (Kierkegaard, 1980, 16-17). Furthermore, we can find an Aristotelian potential in Kierkegaard’s ethics where he writes that the ethical self has chosen ‘himself, then, as an individual who has these talents, these passions, these inclinations, these habits, who is under these influences . . . . [H]ere, then, he has himself as a task, in such a sort that the task is principally to order, cultivate, temper, enkindle, repress, in short, to bring about a proportionality in the soul, a harmony, which is the fruit of the personal virtues (E/O, II, 267).’

Let us then examine Aristotle’s concept of virtue as achievable for humans in order to see how Kierkegaard could have understood moral failure and how to respond to it in an Aristotelian way. He says, ‘Virtue, then, is a state of character concerned with choice, lying in a mean, i.e. the mean relative to us, this being determined by a rational principle, and by that principle by which the man of practical wisdom would determine it (Aristotle, NE, 2, 6).’ In striving to develop the virtues as habits, it is obvious to Aristotle that the individual will make some mistakes in either excess or defect, and then gradually find that mean between excess and defect appropriate to her as an individual. Of course, Aristotle points out that there are some actions and feelings such as murder or jealousy which have no mean (Aristotle, NE, 2, 6). But very little of his ethics is devoted to a consideration of these actions and feelings evil in themselves. Rather the whole thrust of his ethics is to create ways of being happy that are respectful of all citizens, especially through justice, courage, temperance, and wisdom. Moreover, the largest portion of his ethics, namely two books out of ten, is devoted to a rich analysis of friendship as a virtue. He gives the highest praise to and deepest analysis of friendship of virtue. Note this rich analysis by one commentator:

Friendship between A and B is a good for A (an aspect of his flourishing) and a good for B, too. But to be a friend, A must act substantially for B’s good (not his own, A’s) and must value B’s good for the sake of B (rather than for the sake of what he, A, can
get out of the friendship); in other words, A must treat B’s own good as an aspect of his (A’s) good. Yet at the same time, B must value A’s good for the sake of A, and treat A’s good as an aspect of his (B’s) own good (Finnis, 148-149).

When these two friends act for the good of themselves and each other, they have certainly transcended selfishness and have a true compassion for each other, like that of a mother for the child of her womb (Aristotle NE, 9:4). And like a mother, they will put their lives on the line in order to protect the life of their friend, even unto death (Aristotle, NE, 9, 8). The ideals of the best kind of friendship, friendship of virtue, are not an impossible ethics for Aristotle. This ethics is realizable through the joint efforts of two friends who are friends for the sake of realizing virtue in each other, and the best proof of the reality of this kind of friendship is found in the actions of those who would die for their friends. As Aristotle affirms,

It is true of the good man too that he does many acts for the sake of his friends and his country, and if necessary dies for them; for he will throw away both wealth and honours and in general the goods that are objects of competition, gaining for himself nobility; since he would prefer a short period of intense pleasure to a long one of mild enjoyment, a twelvemonth of noble life to many years of humdrum existence, and one great and noble action to many trivial ones. Now those who die for others doubtless attain this result; it is therefore a great prize that they choose for themselves (Aristotle, NE, 9, 8).

Kierkegaardian Ethics

In summary, Kierkegaard has assumed that Christian ethics demands an impossible perfection, but I have identified some scholars’ analyses of a Christian ethics that is not aimed at an impossible perfection, but at a realizable compassion, and I have also developed Kierkegaard’s own suggestion to look at the Aristotelian concept of ethics as rooted in virtue and the realistic possibilities of attaining virtue. Nevertheless, in this next part of the paper I will look at the foundation of ethics for Kierkegaard and see whether or not he can establish a more solid basis for an ethics dedicated to an impossible ideal of perfection.

Walter Sikes evaluates Kierkegaard as offering a freely chosen, existential commitment to the fundamental moral law of the ethical sphere of life, writing:

There is no law of necessity that moves the possible into the actual. The passage from the aesthetic to the ethical sphere of existence is not made by way of some bridge between the two. There is no bridge but instead a great gulf. The transition is always a “leap” that is never devoid of reflection though never produced by reflection. It is an “internal decision in which the individual puts an end to mere
possibility and identifies himself with the content of his thought in order to exist in it.” The leap by which one actualizes existence in the ethical sphere consists of a radical decisive act—a passionate choice. (Sikes, 1968, 57; see also Kierkegaard, 1941, 302; 304, footnote).

Kierkegaard has reinterpreted ‘Know thyself’ as ‘Choose thyself (Kierkegaard, 1959, II, 260).’ The Socratic commitment to the examined way of life is the central ethical commitment in an individual’s life. Kierkegaard has emphasized this commitment as a choice. Hence he has placed the responsibility upon the individual in the ethical sphere for succeeding or failing in his moral life. If the individual does not know himself or if the individual does not do the morally correct action, it is because he has chosen not to know himself and/or not to be morally good (Kierkegaard, 1954, S unto D, 225).

Because Kierkegaard has emphasized choice both in establishing and in achieving the ethical commitment, the individual’s ethical reflection:

... makes the individual realize two things: (1) his distance from the ethical ideal and (2) the need for a free decision to attain the eternal ideal. (1) Ethical reflection dissolves the esthetic dream that the individual man naturally coincides with the conditions of his enduring happiness. (2) Ethical reflection dissolves the esthetic dream that the individual can avoid the strenuous (Collins, 1962, Three Paths in Philosophy).

It is only free choice in the form of a lasting commitment which can free the individual from the aesthetic dream that human fulfillment lies in the momentary and the immediate. It is only free choice in the form of a whole-hearted commitment which can lead the individual to the attainment of the ideal. However, here precisely is where the problem lies for Kierkegaard. The individual by one’s own free effort cannot live up to the moral ideal which one has chosen. Collins explains:

Despite his emphasis upon decisive moral action, the ethical person is likely to become lost in endless reflection upon the unconditional demands of the moral law and the inadequacy of his human resources. This comparison between the moral norm and particular human actions brings the purely ethical life to a standstill—via an “overdose of repentance.” The unique feature of this repentance, however, is that it does not concern past deeds so much as future attempts to achieve the moral ideals (Collins, 1962, Three Paths in Philosophy).

If Kierkegaard is correct that every human individual who attempts to live the ethical commitment fails of necessity to do so, then the ethical commitment leads the individual to despair of necessity. For, if every person must fail in living out one’s ethical commitments, such a person must realize that the self
cannot achieve the moral commandment by oneself and must need to despair and give up the attempt to do so by oneself. The only way in which this despair can be transcended would be through repentance to a personal God who is the source of the moral law and therefore above the moral law and who thereby can forgive the individual for breaking the universal laws. The ethical ideal as a universal ideal works, Kierkegaard says, when the individual lives by that ideal, that is, when he incarnates the universal into his particular existence. However, ‘when the individual by his guilt has gone outside the universal, he can return to it only by virtue of having come as the individual into an absolute relationship with the absolute [namely, God] (Kierkegaard, 1954, F & T, 108) because, Kierkegaard claims, the moral law as a universal ideal cannot apply to the individual who has placed himself outside the universal. The assumption of Kierkegaard certainly appears to be that the ethical category of the universal cannot include forgiveness offered by either other human selves or oneself. The ethical category of the universal needs to be suspended by the creative author of the ethical, God. Kierkegaard seems to be leaning toward towards the view in the *Euthyphro* that God’s command alone creates and/or suspends the ethical good rather than the opposing view that the ethical good is what controls God’s command. So, in Kierkegaard’s view, God command wills that humans attempt the impossible precisely in order that individuals will need to realize their absolute dependence upon God’s will and command. Consequently, ‘as soon as sin makes its appearance ethics comes to grief precisely upon repentance; for repentance [that is, recognition of the self that it has failed and will always fail to live up to the ethical,] is the highest ethical expression, but precisely as such it is the deepest ethical self-contradiction [because the individual out to be ethical but finds it impossible to achieve] (Kierkegaard, 1954, F & T, 108).’ If every individual must fail in living the ethical commitment, then the highest level of subjectivity, the deepest level of intensification of subjectivity, which the ethical can attain to is to persist through the despair of sin to repentance, but such repentance would be the first stage of the ethical religious way of life for such repentance would open oneself to choosing to live completely dependent upon God’s will, the creative source of the ethical.

The assumptions of Kierkegaard which make repentance the next step in the development of the self are: (1) everyone fails to live up to the ethical ideal and (2) the ethical level itself cannot forgive the individual since he has placed himself outside the universal ideal of the ethical. Kierkegaard does not make these assumptions without careful reflection. In the second volume of *Either/Or*, Judge William includes the sermon of an obscure Jutland pastor who ‘has apprehended what I said to you and also what I was desirous of saying (Kierkegaard, 1959, II, 342).’ And the title of the sermon is ‘The Edification Implied in the Thought That as Against God We Are Always in the Wrong (Kierkegaard, 1959, II, 342).’

(1) *Everyone Fails the Ethical*

Mackey explains how even the good person must consider himself guilty, as stated in point (1) above, as failing to live up to the ethical:
To make his point, the priest examines a relationship between two lovers. Love is the Judge’s specialty, but the priest considers a possibility that never arises in the Assessor’s life or thought. Suppose, he says, a radical conflict between lover and beloved. How can such a conflict be reconciled? Will the lover spend himself computing the rights and wrongs of each party to the relationship? Even if that were not an impossible task, such pettiness would hardly bespeak a deep and sincere love. Will he assert his own right against the beloved. If he does, then it is not the beloved but his own rectitude that he loves. The true lover will neither defend himself nor bargain for advantages; the true lover will without reservation put himself in the wrong that his beloved may be right and their love secured (Mackey, 88).

Kierkegaard points out that we relate differently to those we love and those we do not love. When someone we do not love offends us, Kierkegaard says, ‘Your soul is not callous to the suffering they inflict upon you, but search and examine yourself and are convinced that you are in the right (Kierkegaard, 1959, II, 349).’ But suppose you are offended by one whom you love. If you were to think that you were in the right and the beloved in the wrong, ‘this thought would not tranquilize you, you would explore anew every possibility. . . . You would wish that you might be in the wrong, you would try to find something which might speak in his defense, and if you did not find it, you would first find comfort in the thought that you were in the wrong (Kierkegaard, 1959, II, 349–350).’

There is, Kierkegaard points out, no necessity which requires the lover to admit that he is in the wrong. For the evidence is either clear that the beloved has done wrong or ambiguous about who is at fault (Kierkegaard, 1959, II, 349). The lover does not for the sake of truth have to admit that the lover was at fault. There is no logical necessity demanded by evidence which is clear and forceful. But there is the requirement of love that the lover seek to advance the good of the beloved, and for this reason the lover is willing to admit that the fault was his. Now, of course, if the beloved is sincere in her love, then she will also admit that the fault was hers in order to secure their union and acceptance. But then we would have to say that this reconciliation of the two lovers is precisely a choice on their part.

It appears difficult to reconcile the lover’s assumption of guilt with Kierkegaard’s emphasis upon ‘Know thyself’ as ‘Choose thyself.’ If, as Kierkegaard has assumed, the lover was in the right, would we not have to say that the lover is knowing himself falsely when he takes upon himself guilt? If the beloved has free choice in the relationship and is the one who has chosen to do wrong, how is the beloved’s self-knowledge advanced when someone else takes the blame? For example, consider the story of the Prodigal Son. When the son returns home and asks forgiveness for his sins against the father, the father does not say that the son was not guilt; rather, the father simply forgives the son. The story would appear to contradict Kierkegaard’s analysis of the
love which has been wronged.

However, in defense of Kierkegaard we could note that the father may have reflected upon the role he had in his son’s wrong-doing. For it was the father who foolishly gave a rich inheritance to a young son who was not mature enough to use it properly. Could the father really assume that he was innocent in the relationship? Surely the evidence would be ambiguous, and unselfish love would require that the father seek forgiveness from the son just as the son seeks forgiveness from the father. But this need for mutual forgiveness takes us to a discussion of the assumption (2) of Kierkegaard, that the ethical level cannot offer forgiveness to the individual outside the ethical. Can the father and son really forgive each other if Kierkegaard is right? This is the next point to be discussed.

(2) The Ethical Cannot Forgive

The key argument of Kierkegaard for claiming that the ethical cannot forgive significant ethical failure is his assumption that the ethical as the universal cannot apply, that is, forgive, the one who in choice of evil or sin has placed himself outside the universal. As so stated in the abstract, the assumption seems to be true: the universal ideals of morality do not apply to the particular individual who refuses to achieve those ideals. However, we have to remember that the universal ideals of morality include both judgments of moral good and of moral evil. We remember that Judge William himself pointed out that the ethical commitment is a choice of living under the categories of both good and evil (Kierkegaard, 2004, 486). Hence, one who chooses to be ethical but then fails to live up to that commitment can be subsumed under the category of evil-doer or sinner just as much as the one who lives up to one’s ethical commitment can be subsumed under the ethical category of the virtuous person.

Still, we must grant that Kierkegaard’s assumption that an abstract universal ideal of morality cannot forgive the evil-doer is absolutely correct. An abstract ideal cannot forgive anyone, but neither can an abstract ideal honor anyone. The ethical acts of forgiving the evil-doer and of honoring the virtuous person are both acts that must be performed by the individual persons. So Kierkegaard is right that ethical repentance brings the individual into a personal relationship with another person. He claims this other person can only be God, but the other person who forgives could simply be a magnanimous human. For example, if I have offended myself, then in a very real sense I must be another person to myself and summon the courage and compassion to forgive myself. Ethical repentance in this light reveals that truth is subjectivity, that I understand who and what I am and can become only insofar as I become a loving, forgiving person to myself. Also, if I have offended another, that person can discover that truth is subjectivity in that the person has to go beyond the hurt self to the discovery and courageous creation of the loving, forgiving self. Only by developing the resources of one’s subjectivity, only by going beyond rigid, unforgiving ‘superego’ or neurotic moral commands of absolute perfection and by developing flexible ‘ego’ or reasonable moral ideals, can the individual be one’s true self to one’s own self and to others.
An objection has been raised by a reader whose challenge is; ‘Insofar as a cause can be attributed to any “choice of evil or sin”, why is not every act of “of evil or sin” to be approached in the same manner of “unselfish love”?’. My argument for ethical forgiveness of significant ethical failure may have overlooked that there may be different levels even in significant ethical failure, for example, between manslaughter and capital murder. Whereas manslaughter could be ethically forgiven by the victim’s family but at the same time legally punished by the state, say, with imprisonment of one to five years, capital murder, the objection would be, should not be forgiven but should be punished severely, say with life imprisonment with no possibility of parole. However, in answer to that objection, we may note that families of capital murder victims have reached out to the murderers in their prison sentence and have forgiven them for that horrible crime (Kottoor, 1). The thesis of my paper is not that such horrible crimes should be forgiven but that it is existentially possible to choose to forgive such horrible crimes. Such a choice is most difficult, but it is not ethically impossible.

Another reviewer finds an apparent inconsistency in the paper: Kierkegaard is depicted as both committed to forgiveness and blameworthiness, but you never provide a standard for resolving this apparent contradiction.’ In response I would argue that if an action, for example, manslaughter, is not blameworthy, then there is no need for forgiveness. If one person were to kill another accidentally without any fault or guilt, there is no need for forgiveness. The killing was an accident. However, when an action has been judged blameworthy, that moral judgment is appropriate and reasonable. Nevertheless, the judgment of guilt looks to the past of the individual, but in an existentialist ethics rooted in how we choose to respond, oriented to the future, if forgiveness of a blameworthy individual can assist both the one forgiven and the one forgiving build a better moral future, then forgiveness is possible as a reasonable choice, So there is no contradiction between a judgment of guilt and forgiveness because guilt looks to the past of the individual but forgiveness looks to the future.

In summary, it is the thesis of this paper that, if we conceive Kierkegaard’s ethical level of existence as involving the person’s chosen relationship to one’s own self and to others, it is reasonable to conclude that the ethical level can incorporate loving forgiveness as a basic attitude towards self and others which brings about personal growth. A fundamental objection not yet raised could be that such loving forgiveness towards self and others could be a rationalization of my lackadaisical moral failures. Indeed, this choice of loving forgiveness could involve the self in self-deception and lead not to moral growth and development, but to even worse moral failure and degradation; and we must grant that such self-deception could occur. However, if we adopt a Socratic openness to dialogue with others about our moral behavior and fundamental moral dispositions, we open ourselves to evaluation by others who can challenge us to further moral growth. Following Wittgenstein’s argument against the possibility of a private language, we should affirm that our moral language to describe and analyze our actions and our dispositions is not a private language but a language we have learned in a community and which
can be challenged and improved upon for the purpose of moral growth.

In contrast with the thesis of this paper, Kierkegaard holds that moral values are best seen as absolute ideals which demand perfection which humans necessarily fail to achieve, and it is primarily Kierkegaard’s ethics that this paper has evaluated as demanding an impossible perfection which leads the person to despair. This paper has drawn upon Aristotelian resources to affirm that moral ideals are guidelines meant to bring about gradual human development in both the better knowledge of and the increasing realization of those ideals is to give to humanity a morality that certainly allows for people to forgive each other and even themselves as a better way to develop our human potential for morally good behavior.

Bibliography


Rudd, A. J., “Kierkegaard and the Skeptics,” *British Journal for the History of
Kierkegaard and Moral Guilt:
Can the Ethical Forgive Significant Ethical Failure?

In this paper I will discuss the role and meaning of wonder as decisive force in philosophising as obvious in ancient Greek philosophy and in the writings of Ludwig Wittgenstein. I will first discuss the metaphorical approach toward the world in Homer and Hesiod and then the shift from myth to logos as discernible in the Pre-Socratic philosophers, Plato and Aristotle. Finally I intend to convey the dimension of wonder in Wittgenstein’s philosophising – in the way of his regarding both the phenomenal world and the world ‘outside the world of facts’, and in his attitude toward ethics and the limits of language. In his Lecture on Ethics, Wittgenstein mentions the experience of ‘wonder at the existence of the world’ as his ‘experience par excellence’ for his understanding of ethics. I will distinguish between wonder as aesthetic contemplation and wonder as puzzlement and questioning, thus between wonder as silent admiration of the world on the other and a dynamic preoccupation with the manifold and ever-changing aspects of every object of philosophising on the other. Besides, I will show in how far wonder in Wittgenstein’s works can be compared to the ancient Greeks’ thaumazein and to Plato’s and Aristotle’s remarks about wonder as the beginnings of philosophy.

στον Λ.

‘For men were first led to study philosophy, as indeed they are today, by wonder. […]’ (Aristotle, Metaphysics 982b)

Wonder as a decisive force in philosophising – in fact, as its beginning, as Plato and Aristotle put it¹ – has not lost its relevance since. It is indispensable for any serious preoccupation with the complexity of the world and its resulting philosophical problems as it was back then.

Frequently, as Aristotle said, ignorance and the search for knowledge play a part, however, rational explanation all too often can destroy the experience of

¹Cf. Plato, Theaetetus, 155c-d and Aristotle, Metaphysics, A 2, b 982 12.
wonder, as Wittgenstein remarked. (C&V, 5e) In his case, wonder, on the one hand can be observed as an ethical attitude of silent admiration of the world, on the other it leads to questioning philosophical problems in an attitude of never-ending puzzlement at the variety of aspects.

**Wonder in Ancient Greece**

In order to consider the dimension of wonder in ancient thought, it is necessary to begin with men of thought and poetry like Homer and Hesiod. Between them and the following philosophers both parallels and decisive differences can be observed, one of them in their attitudes toward religion. Whereas in Homer man had an anthropomorphic idea of gods while searching to explain the world by metaphors and myths, the first philosophers – the so-called Pre-Socratic thinkers – were striving for rational and scientific explanation. Still, mythological elements continued to pursue – as can partly be seen in Plato.

The aspect of wonder in Homer, Solon, Hesiod – and in part in the Ionian Natural philosophers – can above all be defined as an act of admiration, also ‘astonishment’ in the sense of puzzlement, a curiosity of the cosmos. (Martens 2003, 20) The wish to explain the mysterious is satisfied by mythologizing the world of the gods. The typical idiom for looking in wonder – the *thaumazein* [*θαυμάζειν*] – occurs in Homer as *thauma idesthai* [*θαύμα ιδέσθαι*] meaning something like ‘to look at a miracle’. Apart from everything transcending the phenomenal world, it was above all beauty and greatness which evoked wonder in man – Richard Kroner sees beauty as the key of understanding the spirit of ancient Greek thought and refers to Hegel, who described the religion of the Greeks as ‘religion of beauty’. Even though the Greeks officially admired the gods, secretly and implicitly they admired the cosmos as the most divine. (Kroner 1957, 46). This unconscious and unspoken pantheism hidden in polytheism was a presupposition for religion and art in ancient Greece. (Cf. ibid.)

The Greeks’ admiration of the cosmos with its mysterious phenomena in nature thus expressed itself in stories, even though rational thinking gradually developed side by side with a mythical apprehension of the world. Hesiod for the first time poses the question for the creation of the world, his cosmology is a synthesis of cosmogony and theology (Kroner 1957, 72). However, he remains captured in epics, thus creating his *Theogony* from earlier myths.

In ascribing to the muses the ability to find the unhidden truth (*aletheia* [*αλήθεια*]) – and not only the illusory half truths as in Homer’s works – Hesiod took a first step beyond myth toward philosophising by searching for utter truth and thus knowledge. According to Schadewaldt, this was one of the beginnings of philosophy. (Schadewaldt 1978, 85) Similarly, Snell sees in Hesiod’s *Theogony* a ‘remarkable step on the way from epics to philosophy’ und thus
calls Hesiod due to his questions after the origin and nature of the world the “predecessor of philosophy.”” (Snell 1993, 20)

In his genealogy of the gods, Hesiod describes *Thaumas* [θαύμας, root: θαύμαντος: wonder] as the child of *Pontos* (God of the sea) and of *Gaia* (Goddess of the earth). The fact that *Thaumas* viz. ‘to wonder’ is described as a child of the sea, can be understood by the close relationship of the Greeks to the sea which appeared to them as a sphere of miracles, constantly changing in light and colour.

In addition to Homer’s anthropomorphic idea of gods, Hesiod sees them also in natural phenomena such as the sun, the sea and the moon, and in trees and rivers emerging there as titans, Erinyes and Muses.

Thales of Miletus and the subsequent Pre-Socratic Philosophers, who endeavoured to explain the world by scientific methods instead of myths went one step further. In reinterpreting the God *Okeanos* as the general principle of water, Thales, according to Aristotle, took the step from myth to logos. The succeeding philosophers such as Anaximandros, Anaximenes were still striving at explaining the principle of the world – either by air, fire or *apeiron* (the infinite, indefinite). Heraclitus and Parmenides rejected a merely sensory perception of the world and emphasized rational thinking as path to knowledge and truth. The silent unquestioned admiration of the cosmos gradually moved toward a sense of puzzlement which later in Plato and Aristotle became a source of tension. In their philosophies the former reflections about the cosmos turn to reflections about man and his existence in the world.

In this sense, Socrates was concerned about man and his search for truth. In his method of elenchus he tried to lead his pupils to philosophical arguments in order to sharpen their sense for truth. While questioning everything, he leads them to ever new perplexity and astonishment – not unlike Wittgenstein in his *Philosophical Investigations* as we shall see later on. Socrates’ method of dialectics of specific questioning and answering finally returning to its start – the non-knowledge – can to some extent be compared to Wittgenstein’s attitude of wonder at essential matters that cannot be explained.

In his well-known remark about wonder to Theaetetus, Plato alludes to Hesiod’s picture of Iris, the daughter of *Thaumas*:

> For nothing else is the beginning (principle) of philosophy than this, and, seemingly, whoever’s genealogy it was, that Iris was the offspring of Thaumas (wonder) it’s not a bad one. (*Theaetetus* 155c-d)

The picture of the rainbow which stimulates to wonder at the colours emerging from sun and rain and which is regarded as the daughter of *Thaumas* – the personified wonder – is characteristic of Greek mythology with its colourful metaphors. The unexplainable is depicted in myths in order to render what cannot be grasped by ratio, but only shown, as Wittgenstein would say.
Despite a critical attitude toward myths, Plato still makes use of metaphors and similes, sometimes drawing back to myths. In the dialogue *Phaedrus*, e.g., myth, apart from dialectic speech, serves as means of education, and the myth of cicadas is a wonderful example of oblivious dedication to the beauties of the cosmos leading to creativity as illustrated in the cicadas which in their devotion to endless singing forget about any physical needs and therefore stand for the ‘prophetesses of the Muses’.

For Plato, wonder is a necessary step for philosophising, a pre-requisite for the man who would love wisdom – the *philosophos* – and thus starting point for knowledge. In the *Symposion*, in the discussion between Socrates and Diotima on the nature of Eros, the word *thaumazein* occurs several times – both as admiring wonder and puzzlement about what Diotima tells Socrates about love and the beautiful – i.e. the act of philosophising proceeding from love of sensory beauty to the beauty of souls and ultimately epistemological truth (210e) – ‘the good and the beautiful’, the ‘idea of ideas’, the divine good itself. Plato’s high valuation of the beautiful is reminiscent of Homer’s admiration of the cosmos and the divine, but in a new, reflected and abstract way. In Plato the sensory is transformed into the sublime so to speak. Besides, the ethical component expressed by *kalokagathia* [καλοκαγαθία] is of central relevance.

In Aristotle we find the relationship between silent wonder and the emergence of scientific questions arising from astonishment.

For men were first led to study philosophy, as indeed they are today, by wonder. At first they felt wonder about the more superficial problems; afterward they advanced gradually by perplexing themselves over greater difficulties; e.g. the behaviour of the moon, the phenomena of the sun, and the origination of the universe. Now, who is perplexed and wonders believes himself to be ignorant. (Hence, even the lover of the myths is, in a sense, a philosopher, for a myth is a tissue of wonders.) Thus if they took to philosophy to escape ignorance, it is patent that they were pursuing science for the sake of knowledge itself, and not for any utilitarian applications. (Aristotle, *Metaphysics* 982b).

Thus Aristotle, too, refers to the first men of thought in ancient Greece and emphasizes the significance of both myths and the search for knowledge, while refusing any practical purpose. Yet, despite his appreciation of myths, Aristotle’s words reveal a preference for a rational view of the world. At that time, *theoria* as originating in an attitude of wonder was seen as characteristic of the philosopher in order to perceive the truth.

Whereas wonder as a contemplative attitude lacks rational considerations, these are necessary for wonder as puzzlement in order to deal with epistemological questions. In the following I will discuss in which sense wonder in Wittgenstein can be seen as an attitude of contemplation or as a state
of puzzlement, i.e. which elements of intellectual perception – intuition or ratio – prevail.

Wittgenstein and Wonder

In his manuscripts, occasionally Wittgenstein speaks of ‘wonder’ – in connection with his understanding of ethics, his criticism of science and his preoccupation with aspect-seeing in the Philosophical Investigations. However, the term wonder does not occur often and it is never the object of conceptual analysis. This fact is quite telling: for the person who wonders is so sensitive that ‘he rather suffers of any language, one could say: he ought to be mute if language had not been invented by others.’ (Bloch 1996, 16)

Apart from a few explicit remarks about wonder, Wittgenstein’s approach toward the objects of his philosophising – be they phenomena of the visible world, fictitious examples he simply imagined, or questions concerning the realm of the invisible – can be said to be characterized by an attitude of wonder, this in both an ethical and aesthetic sense, while both aspects cannot really be separated from one another but must be seen as complementary, thus consistent with Wittgenstein's remark ‘Ethics and Aesthetics are One’. (TLP, 6.421)

Notebooks 1914-1916 and Tractatus: Wonder as Aesthetic Contemplation

As regards the aesthetic significance of wonder, we are reminded of the aesthetic contemplation described by Schopenhauer – a contemplation in which the ‘pure subject of knowledge’ is absorbed by looking at an object, so that subject and object become one. The subject is elevated beyond the world of suffering, and – on finding himself on a higher level – intuits the ‘Platonic idea’ in the object of his contemplation. In these ‘rare happy moments’ any analysing form of questioning and attempts at explanations would be out of place. In his apprehension of the eternal and universal idea inherent in the transitory, concrete object of contemplation (be it a tree or an object of art) the aesthetic observer consists merely of intellect, deprived of all forms of the principle of sufficient reason, i.e. of space and time. There is a strong ethical component in Schopenhauer's description of the aesthetic contemplation – namely in so far as the observer's intellect rises beyond the sensory to an extent that he is spiritual being.

As to Wittgenstein, there are parallels to Schopenhauer as early as in the Notebooks 1914-1916 and in the Tractatus. This applies above all to the passages where we can speak of mystic-pantheistic tendencies in Wittgenstein and of his ethical and aesthetic considerations. Wittgenstein emphasizes the

---

1Whereas Plato differentiates between phenomenon and idea, Schopenhauer distinguishes between the will as thing in itself (in a metaphysical sense) and the will as adequate and inadequate manifestations in the world. Schopenhauer’s idea is merely the adequate manifestation of the will in contrast to the inadequate manifestations represented by the individual, transitory phenomena.
The work of art is the object seen sub specie aeternitatis; and the good life is the world seen sub specie aeternitatis. This is the connexion between art and ethics.

The usual way of looking at things sees objects as it were from the midst of them, the view sub specie aeternitatis from outside. In such a way that they have the whole world as background. Is this it perhaps – in this view the object is seen together with space and time instead of in space and time? (NB. October 7, 1916)

Wittgenstein’s use of the concept sub specie aeternitatis denotes a view of the world which transcends the world of facts. Whereas the so-called ordinary and scientific view of things follows the principle of sufficient reason and is thus restricted by the forms of time and space, the view sub specie aeternitatis transcends time and space. In the Tractatus, 5.61 Wittgenstein writes that logic pervades the world and this does not only mean that the limits of the world are the limits of logic, but it suggests that in the world nothing happens which would contradict the laws of logic viz. the laws of God’s eternal nature in a Spinoza’s sense. In a view sub specie aeternitatis, things are perceived in their necessary and logical connection. At the same time one recognizes that the facts of the world are not the end of the matter, but that what is essential – the meaning of the world – lies outside space and time.

‘To view the world sub specie aeterni is to view it as a whole – a limited whole. Feeling the world as a limited whole – it is this that is mystical.’ (TLP, 6.45)

To be happy means to devote one’s life to the spirit, freed from the desires of the will, and in Schopenhauer’s terms to merely contemplate, as a ‘clear eye of the world’. (WWR I, § 36, 186). ‘Is this the sense of the artistic way of looking at things that it looks at the world with a happy eye?’ Wittgenstein asks (NB, October 20, 1916). What in the Tractatus he defines as the mystical, in the Lecture on Ethics as wonder at the existence of the world, is here described as aesthetic miracle: ‘Aesthetically, the miracle is that the world exists. That there is what there is.’ (ibid.)

In addition to art, Wittgenstein also considers philosophy to view the world from the right perspective: Similarly to the artist who renders something ordinary in a way that it appears as a work of art, thus enabling us to view the things in the right perspective, philosophy is capable of doing this, as well:

‘But now it seems to me too that besides the work of the artist there is another through which the world may be captured sub specie aeterni. It is – as I believe – the way of thought which as it were flies above the world and leaves it the way it is, contemplating it from above in its flight. (C&V, 7)

Whereas the view from above reminds us of the mystical and somewhat stoic attitude toward the world as discernible in the Tractatus, the reference to
‘thought in flight’, though, suggests the dynamic process of movement characterizing Wittgenstein’s philosophising in later years. The demand to leave the world the way it is further suggests a distancing oneself from any intrusion akin to scientific explanation.

The Lecture on Ethics

In this lecture Wittgenstein ventured to convey of what he considered the essence of ethics. As he was against any objective theory of ethics in philosophy, he spoke from his subjective point of view, giving three examples of his personal experience while treating problems of ethics in connection with problems of language and differentiating between absolute and relative value. ‘Wonder at the existence of the world’ – his ‘experience par excellence’ – reveals its ethical and absolute value in an attitude of aesthetic contemplation and perceiving the world as miraculous and not self-evident. In everyday language this may sound nonsensical – in contrast to wonder at something extraordinary never seen before that would evoke astonishment to everyone. The same applies to the feeling of absolute safety: Whereas the sentence ‘I feel safe in the house when it is raining outside’ is understandable in everyday language and thus to be considered a meaningful sentence, though of relative value, the sentence ‘I feel safe whatever happens around me’ in an absolute sense would strike us as nonsense. Here we come to the border between meaningful and nonsensical expressions – between the sayable restricted to facts and the unsayable or ethical concerning a ‘higher’ sphere. As already hinted at in the Tractatus, there are no ethical propositions, because propositions cannot express something higher. (TLP, 6.42) Therefore, the characteristic of any ethical and religious expressions is their nonsensicality. Every attempt at expressing in words something about ethical values only reveals the limits of language and would thus be a ‘running up against the walls of our cage’. For Wittgenstein, ethical propositions are rooted in a superior – a divine – authority which is unassailable and therefore beyond any explanation. ‘What is ethical cannot be taught. If I could explain the essence of the ethical only by means of theory, then what is ethical would be of no value whatsoever.’ (WVC, 117).

In none of Wittgenstein’s other written works is the connection between wonder, ethics, aesthetics and language in the sense of being reduced to silence as obvious as in the Lecture on Ethics.

Philosophical Investigations: Wonder as Puzzlement and Questioning

Wittgenstein's attitude of wonder cannot only be observed under ethical and aesthetic aspects viz. in aesthetic contemplation, but also in his philosophical method, i.e. in his philosophical dialogues which aim at investigating every object from various perspectives. Insofar Wittgenstein did not maintain a silent and stoic attitude toward the ‘existence of the world’, but his philosophising leads to movement, to creativity, to a ‘change of aspects’. Yet, the perception of a change of aspects is not possible without an attitude of
wonder. His incessant way of writing several variants of formulations of a sentence hints at how he tried to present his philosophical concern – his awareness of subtle nuances in the meaning of our thoughts, our different views at things and thus manifold ways of interpreting the phenomenal world. Thus, the dimension of wonder in Wittgenstein is closely connected with the way we see and interpret things. Just as the meaning of words changes according to their occurrence in different contexts, any objects, pictures and illustrations etc. can be seen and interpreted from various perspectives. Since the world in itself is continually flowing – as Heraclitus emphasized –, it is important to recognize as many perspectives as possible in order to take them into consideration when we want to describe them. The perception of different aspects is facilitated by the ability to wonder, but it is accompanied by ratio: ‘I think it could also be put this way: Astonishment is essential to a change of aspect. And astonishment is thinking.’ (LWPP I, § 565)

‘What is strange is really the surprise; the question: “How is it possible!” It might be expressed by: “the same – and yet not the same.”’ (LWPP I, § 174)

The decisive point is the change of aspects, i.e. the moment when we perceive the change. This moment evokes wonder: ‘But the change evokes the surprise not produced by the recognition.’ (LWPP I, § 517)

This is but a short moment in which we see how the new aspect ‘dawns’ – a moment which is accompanied by a state of wonder within us. ‘The expression of seeing an aspect is the expression of a new perception.’ (LWPP I, § 518) One gets the impression that the picture or object changes, yet it is only the impression that changes, the view of the observer – his way of perception.

Wittgenstein’s Critical Attitude toward Science

‘Man has to awaken to wonder – and so perhaps do peoples. Science is a way to sending him to sleep again.’ (C&V, p. 5e)

Wittgenstein’s sceptical attitude toward progress, technology and science can be observed in many passages, just as his ambivalent attitude toward analytical methods in philosophy:

The mathematician too can of course marvel at the miracles (the crystal) of nature; but can he do it, once a problem has arisen about what he sees? Is it really possible as long as the object he finds awe-inspiring or gazes at with awe is shrouded in a philosophical fog? I could image someone admiring trees, & also the shadows, or reflections of trees, which he mistakes for trees. But if he should once tell himself that these are not after all trees & if it becomes a problem for him what they are, or what relation they have to trees, then his admiration will have suffered a rapture, that will now need healing. (C&V, 65)

The crucial point here is to free oneself of a scientific as well as philosophical attitude in the sense of striving for explanation, in order to
Thaumazein in Ancient Greek Philosophy and Wonder in the Writings of Ludwig Wittgenstein

restore a natural view of the phenomenal world. In an attitude directed toward analysing, the ability to wonder is ‘put to sleep’, or destroyed.

However, it would be misleading to conclude that scientific explanation destroys the miraculous in every case. In his critical remarks on Frazer’s *Golden Bough*, Wittgenstein maintains that every phenomenon can appear mysterious and meaningful to us, not only to the so-called primitive man who has no scientific explanation for, say, phenomena in nature like fire, thunderstorm and the like. (Cf. MS 110, 197f.) In fact, the awakening of the human intellect ought to sharpen our sensibility for what is miraculous and therefore worth being met in an attitude of wonder. It is in this respect that one could refer to ancient Greek philosophers who gradually moved from a silent admiration of the cosmos to a striving after explaining in a scientific way. Still, they did not lose their respect toward the miraculous of the world. *Thaumazein* is a means to reveal the mysterious in order to achieve *aletheia* and thus philosophical truth.

The Ethical Aspect of Wonder

Apart from the *Lecture on Ethics*, the ethical aspect of wonder is particularly discernible toward the end of the *Tractatus*, where Wittgenstein holds what one could describe as a mystical attitude toward the world: ‘It is not *how* things are in the world that is mystical, but *that* it exists.’ (TLP, 6.44)

Wonder at the ‘existence of the world’ viz. the world ‘*that* it exists’ incorporates an ethical dimension insofar as wonder at something apparently self-evident such as the world presupposes to be conscious of its preciousness and uniqueness. It requires a different kind of view than the normal one – a view in the light of eternity and not in time and an attitude of awe toward the realm beyond which eludes rational explanation. As Pascal writes about the position of man in nature, man would ‘shudder’ at the sight of the wonders surrounding him, and he would rather be prepared to contemplate in silence but striving to explore them. (Pascal 1956, 31)

Wittgenstein’s wonder at the existence of the world lies on the same level as his attitude of silence toward the sphere of the ineffable: in other words, the ethically grounded wonder leads to an ethically grounded silence. As he remarked to members of the *Vienna Circle*: ‘Astonishment at the fact of the world. Any attempt to express it leads to nonsense.’ (WVC, p. 93) As discussed before, wonder is closely connected with ‘nonsense’ which Wittgenstein was aware of in any attempt at verbalizing ethical and religious matters, for this would go beyond the limits of meaningful language and, at the same time, beyond the limits of the world. Consequently, he chose to keep silent instead of ‘babbling’ about questions that cannot be answered sufficiently in the world of facts, as the ‘meaning of the world must lie outside the world’ (TLP, 6.41)

Wittgenstein’s philosophising has a strong ethical component that can be perceived even in his style of writing. There is a kind of subtlety that finds expression in an extraordinary sensibility for the language he uses: ‘O, why do
I feel as if I wrote a poem when writing philosophy? It is as if there was something small that has a wonderful meaning. Like a leaf, or a flower.’ (MS 133, 13r)

The more precious and ‘higher’ the object of philosophising appears to Wittgenstein, the more modest he becomes – being conscious of the limits of language. Thus the distinction between saying and showing as discernible in the *Tractatus* he basically kept throughout his writings. What can be *shown*, can be experienced by the ability to wonder and at best be expressed aesthetically, e.g. in poetry, painting or music. Wittgenstein's frequent use of metaphors, similes and fictitious examples is a means to at least hint at what cannot be expressed by ordinary language. His respect for both the visible world and the world beyond is connected with his awareness and acknowledgement of the limits of scientific-rational explanations.

**Conclusion**

Even though wonder represents the starting point and pre-condition for any kind of philosophising, the experience of speechlessness involved causes a paradoxical state for the person who philosophises, for wonder on the one hand leads to a deeper understanding of the mysterious and ineffable, on the other hand, though, it eludes conceptual determination and thus the possibility for positive philosophical results. As such, wonder, despite or perhaps precisely because of philosophical insight, leads to a state of failure.

There is no escape but to accept this fact and resign in silence. Wittgenstein must have considered Socrates’ claim ‘I know that I don’t know’ (the knowledge of ultimate non-knowledge) to be a painful paradox – a problem he was unable to solve and toward which he held a controversial position: On the one hand, his analytical investigations into philosophy were of a scientific nature, on the other, he held a distanced and even critical attitude toward rational explanation – an approach in part similar to that of an artist or mystic. Wittgenstein’s two differing positions correspond to his distinction between the world of facts and the sphere outside the world of facts: whereas he considers the former to be grasped in terms of language and of science, the sphere beyond the world of facts – in traditional philosophy usually treated as metaphysics –, to his mind cannot be explained satisfactorily by science. Wittgenstein’s attitude of wonder (and his apparently controversial utterances about the phenomenon) correspond to this dualistic view of the world:

a) Wonder, on the one hand means an ethical-aesthetic contemplation directed toward wholeness and against any form of analysis. This attitude is akin to the ancient Greeks’ silent admiration of the beauties of the cosmos as well as to their ethical approach toward the unexplored, also to the form of *theoria* still described by Plato and Aristotle.
b) On the other hand, wonder means active and discursive questioning, and involves thought, reasoning and investigation in order to restore puzzlement and irritation. It originates in ratio and corresponds to the philosophical urge viz. to the philosophical wonder observable not only in ancient, especially attic philosophy, but later in traditional philosophy, as well.

Whereas the experience of the mystical attitude of wonder cannot be shared via language, the attitude of dynamic astonishment at the individual phenomena of the visible world prepares the ground for a dialogic interaction with a 'you' – real or fictitious – as can be observed in Wittgenstein’s mature philosophy. But even here there is a danger of aporia as Wittgenstein implicitly hints at in the example of the fly-bottle (PI, § 309), where he alludes to the painful, endless process of philosophical arguments from question to answer and then to next question: the never-ending astonishment at ever new aspects of perception in dealing with philosophical problems. The ‘vanishing’ of the problems Wittgenstein did not expect by arriving at solutions, but by achieving ‘dissolution’, not by theory but by activity. In other words, by a radical change in philosophy – enabled by a philosophical method not unlike to a Socratic therapy which, so it seems to me, took its starting point in an attitude of wonder, while at the same time directing individuals to an attentive view of and care for both what is in the immediate surroundings and to an attitude of respect and awe for what escapes philosophical explanation.

References


Part B: Philosophy of Language
CHAPTER EIGHT

Semantic Intentionality and Intending to Act

Dale Jacquette, Universität Bern, Switzerland

The idea that the intentions involved in meaning and action are nominally designated alike as intentionalities suggests, but does not guarantee, any special logical or conceptual connections between the intentionality of referential thoughts and thought-expressive speech acts with the intentionality of action or doing. The present essay asks whether there is a deeper systematic connection between semantic intentionality and intending to act. Four possibilities are considered by which meaning and action intentionalities might be related to be systematically investigated. Meaning-serving and action-serving intentionalities might exclude one another, partially overlap, or partially fail to overlap with one another, or to subsume one in the other in either of two inclusion relations. The theoretical separation of the two ostensible categories of intendings is criticized, as is their partial overlap, in light of the proposal that thinking and artistic and symbolic expression are activities that favor the inclusion of paradigm meaning-serving intentions as among a larger domain of action-serving intentions. The only remaining alternative, including action-serving intentions reductively in meaning-serving intentions, is briefly sketched and defended as offering in an unexpected way the most cogent universal reductive ontology, in which the intentionality of doing generally relates to the specific intentionality of referring in thought intrinsically to the objects of predications, and derivatively with respect to a thought’s artistic and symbolic expression in language, art, and artifacts.

Intentionality of Thought and Action

There appears to be a difference between the intentionality of reference and the intentionality of action. When I act, I intend to do something, but, in the process, it seems at least superficially that I do not always refer to any intended object. If true, this apparent asymmetry minimally suggests that intending to do and intending to refer are so unrelated that an agent can do things without referring to any intended object of thought, even though, in an obvious sense, to refer to an intended object in thought or language is also mentally or linguistically to do something.
The disconnection between intentionality in reference and intentionality in action is sharply pronounced in John R. Searle’s 1983 study, *Intentionality: An Essay in the Philosophy of Mind*. There Searle maintains that to speak univocally of referential intending and intending to act, as collecting under one terminology the intentionality of thought, expression, and action, is at most only a ‘pun’ or equivocation, involving the same word with easily confused but conceptually radically discordant meanings. Searle writes:

...intending and intentions are just one form of Intentionality among others, they have no special status. The obvious pun on “Intentionality” and “intention” suggests that intentions in the ordinary sense have some special role in the theory of Intentionality; but on my account intending to do something is just one form of Intentionality along with belief, hope, fear, desire, and lots of others; and I do not mean to suggest that because, for example, beliefs are Intentional they somehow contain the notion of intention or they intend something or someone who has a belief must thereby intend to do something about it. In order to keep this distinction completely clear I will capitalize the technical sense of “Intentional” and “Intention-ality”. Intentionality is directedness; intending to do something is just one kind of Intentionality among others.

On Searle’s view, it is as though we use the words ‘intend’, ‘intend-ing’, ‘intentionality’ and cognates only through a series of unfortunate linguistic accidents in semantics, philosophy of mind, and action theory. The situation in these philosophical subdisciplines might be comparable, then, to the happenstances in the evolution of colloquial language by which we have come to use the same word ‘times’ for the mathematics of arithmetical multiplication, and for moments of temporal succession. Thereafter, Searle adds, on the same page: ‘Related to the pun on “intentional” and “Intentional” are some other common confusions’.

Such equivocations are rife in ordinary discourse, from which they also contaminate philosophical terminologies. They are taken over from these sources as are all technical nomenclatures, refined for more precise expression in philosophy and other disciplines. We must confront conceptually confused nomenclature that often constrains thought, even if we try to turn away from all those accidental equivocations in the evolution of natural language, and consider instead only an ideal language, such as Gottlob Frege’s *Begriffschrift*. The latter formalism avoids these problems of colloquial expression by syntax restrictions that require all terms in the technical language to have both

---

1Searle (1983), 3.
2Ibid.
3Wittgenstein (1968), §109: ‘Philosophy is a battle against the bewitchment of our intelligence by means of language’.
Sinn and Bedeutung, and preclude any term from equivocally having more than one Sinn, and, hence, from having more than one Bedeutung.¹

Is the situation much the same with respect to the common word ‘intention’, as it is found in everyday and some contemporary philosophical discourse? Are the words ‘intend’, ‘intention’, and ‘intentionality’ used equivocally in speaking of referential semantics and the metaphysics of action? Are the intentionality of reference and the intentionality of action, meaning-serving and action-serving, two distinct and philosophically unrelated things, that historically just happen accidentally to have shared the same form of letters as their common name, but between which respective concepts there is no further interesting conceptual interconnection? Or is there some deeper linkage between the intentions of reference and the intentions of action, suggesting a more unified theory of the two phenomena? If so, then both of these applications of ‘intention’ together might be required in some sense of what is usually considered definitive of a rational responsible human thinker and agent. Does the intentionality of reference and action present us only with a pun, as Searle contends, a bald equivocation, or does the terminology reflect even in its ordinary language roots that conceptually these two categories are somehow inseparably joined?

**Referring as Doing in Thought and Expression**

There are four possibilities in a re-examination of the question as to whether or not there is anything in common between the intentionality of reference and the intentionality of action, keeping physical action or bodily activity especially in focus. We begin by distinguishing semantic and action theory intentionality hypothetically in these ways, allowing that in principle meaning-serving and action-serving senses of intentionality might be:

1. Totally distinct.
2. Partially distinct and partially overlapping.
3. Subsumption of the intentionality of reference by the intentionality of action.
4. Subsumption of the intentionality of action by the intentionality of reference.

We critically consider each of these four possibilities in turn, and in some instances under a variety of plausible branching interpretations.

**Semantic-Action Intending Relation (I)**

Possibility (1) seems to be the position that Searle endorses. It might seem eminently reasonable, particularly in the absence of a plausible theory relating

the two types or applications of intentionality as more than equivocal talk about mutually exclusive concepts, extensionally to divide the intentionality of action from the intentionality of meaning, without so much as a grudging partial overlap.

If we choose to do so, then we gain the further advantage of having distinct concepts of intentionality, each tailored to its task of explaining the directedness of action to a purpose, end or goal, and the relation by which the mind refers to things as objects of predicative and inferential thought, as in a thought’s artistic and symbolic expression and communication. The tradeoff of having distinct specific concepts of meaning-serving and action-serving intentionality, the cost of maintaining such an opulent explanatory parallelism, is that we sacrifice the explanatory aesthetics, generality and elegance of a reductive theory that somehow satisfactorily relates meaning-serving and action-serving intentionality under a single concept of intentionality, of which an intelligent thinking subject and acting agent can have intentions that are relevant in different ways to the intendings involved respectively in meaning and action.

We can test some of these proposals by including nonhuman animals also in our deliberations. Even here, however, we need not regard what animals do as actions in the requisite sense, and we can accordingly understand animal movement as non-intentional or extra-intentional, without worrying excessively about whether nonhuman animals are capable of any sort of semantic intentionality, of referring to things in thought, assuming that they think, or as having evolved some kind of proto-language or primitive entirely extra-linguistic but still symbolic mode of thinking, or even expressive communication.¹

If, however, contrary to Searle, in the human paradigm, that, for obviously and defeasibly justifiable reasons, is taken as the benchmark for intentionality in both semantics and action theory, regardless of how these are logically or conceptually related, we can intend to refer, just as we can intend in other ways to act. If referring is a special kind of doing, a unique category of mental or more overt physical action, typified by speech acts, then Searle cannot be right to draw a sharp distinction between the intentionality of meaning in particular and the intentionality of doing more generally. For I can try and fail to refer, just as I can try and fail to cross the street or tie my shoelaces. I may want to refer to the first Holy Roman Emperor by his historic personal name, but I cannot recall that it is Ludwig I, or I refer to someone else by mistakenly using another individual’s name. Or I may stammer and stutter, in the event, and do not succeed in completing the intended and desired semantic action at all, perhaps not even in thought. Nor should we assume that Searle would want to take issue with any of these commonplace observations.²

¹Condillac (2001).
²Searle in the cluster description theory of reference in his early (1958) essay, is committed to the possibility that I can intend to refer to Aristotle by using the proper name ‘Aristotle’, but fail to do so if I do not have available for explication at least one description that truly applies to the named object. See the essays collected by Tsohatzidis (2007).
The question is whether these shared characteristics, as kinds of intentionality or Intentionality, including properties and especially relations involving psychological occurrences, whether in action or semantic reference, reflect anything essential about the intentionality of reference and the intentionality of action. I think they do, whereas Searle tries to own the word ‘Intentionality’ for semantic purposes, and otherwise considers philosophical discourse about meaning-serving and action-serving intentionality to be at best an amusing equivocation and otherwise at worst a potential source of philosophical confusion.

Searle nevertheless agrees that the two categories of intentionality are properties of thought, of psychological occurrences, although they need not be in any other sense cognitive, for all that Searle argues. They could be abstract, logical, mathematical or metaphysical. Still, they are in all events conceptual, applying in the two cases to a thinker’s psychology. That is already something more than mere random equivocation of the sort by which the same word ‘bank’ is used to refer equivocally both to a financial institution and the edge of a body of water. We do not seem to be dealing with quite so rough a pun, even if Searle is right, but also admittedly with something less than straightforward logical or conceptual association. From the fact that action-serving and meaning-serving intentions are designated alike as involving intentionality, it does not follow that within the general concept of intentionality there is any deeper interrelation between action-serving and meaning-serving intentions. They can belong side by side as completely distinct instances of intentionality, psychological in the most general sense, without there existing any further connection between the intentionality of thought in meaning and in actions, intending, often in deciding, to act. That, anyway, is how one might reasonably attempt to understand Searle in the pun passage from *Intentionality*.¹

If we needed only an overlapping of the properties of the intentionality of reference and of action in order to evidence an overlapping of the intentionality of reference and of action themselves, then it should suffice merely to observe that both share in common being referred to by the same word ‘intentionality’ or ‘Intentionality’¹. It is noteworthy in this regard that semantic philosophy has widely and emphatically recognized that a pure formalism in the theory of meaning cannot hope to attain adequate explication of a word’s or sentence’s meaning without either an integrated or superadded pragmatics, and even a contextualized pragmatics, by which the content as well as the form of symbolic meaning can be represented, and by means of which formally extensionally indistinguishable ambiguities can be properly disambiguated. Later Wittgensteinians should join the chorus against Searle here, because in the posthumous publications, especially *The Blue Book* and *Philosophical Investigations*, Wittgenstein understands the naming of an individual significantly for present purposes as a rule-governed move in a language game, and hence as an action or something that we do.²

²Austin (1975).
We name and use names as tools for a variety of actions we perform in a
greater context of doings. These doings can fit together in complex ways with
more encompassing scopes and in means-ends relations, as when we mix
pigments in preparing to paint, and as potentially part of yet another action to
which the action of mixing pigments may be further subordinate. The same in-
sight is enshrined in the semantic discussion and even in the title of J.L.
Austin’s (1962) posthumously published 1955 William James lectures on How
to Do Things With Words.\(^1\) We do something, accomplish some part of our
business in the world, when we name things or learn their pre-existent names,
and we use the names we devise or acquire in intending to refer to exactly
these things and communicate our thoughts about them with other intelligent
language users for a variety of cooperative and competitive purposes. When we
use propositions in their logical interrelations to construct and critically
evaluate theories, to understand the state of things, and make action plans for
future decision-making and execution. Intending is already doing, on such a
conception, whether it is referring, itself an action, or doing something else
again, engaging in more overt physical action that is itself also at least partly
semantic, or otherwise entirely extra-semantic.

We have in any case by now reached the point where we have cast
sufficient doubt on the truth and perhaps even the cogency of Searle’s charge
that we merely pun when we try to relate the intentionality of meaning and the
intentionality of action as anything more than mutually exclusive. There is
something more interesting going on in the intentionality of referring and
intending to act, given that referring is a kind of mental or expressive act, than
a regrettable historically accidental terminological coincidence. A correct com-
prehensive philosophy of mind and language needs to disambiguate and
somehow relate meaning-serving and action-serving intentionality, if it is going
to make worthwhile progress either in semantic philosophy or action theory.\(^2\)

Semantic-Action Intending Relation (2)

Choice (2) seems to offer a sensible compromise between (1) and some of
the objections that have been raised against it. Perhaps some, though not all,
semantic intentionality is also actional (praxeological) or action-serving. Argu-
ably, some but not all semantic occurrences involve actions.

If I am totally immobile, not doing anything that would obviously
constitute action, I might still be daydreaming, and in my thoughts thinking
about and hence intending the Eiffel Tower, semantically intending the cast-
iron structure, without intending to do anything physical involving the Eiffel
Tower, as I might if I alternatively intend to climb it, or paint it, or blow it up,
or use it as a philosophical illustration. Whereas some semantic intendings are
also action-serving or doings in the ordinary sense, it is not clear that all are.
This intuition suggests support for at least some of their partial extensional
overlap. The second aspect of partial overlap here might then be carried by

\(^1\)Austin (1975).

such examples as those in which actions are generally intentional (unsurprisingly) in the action-serving sense, but not generally intentional in the semantic or meaning-serving sense.

We might find this in cases where we act intentionally (to speak redundantly for emphasis, on the assumption that all action worthy of the name is intentional in the action-serving sense), but where reference or semantic relation is not conspicuously indispensibly involved. For example, I intend to tie my shoelaces, and I do so, although without entertaining any thoughts that intend any particular objects. I do not, at least on some occasions, think about my shoes or their laces or the fact that they stand in need of being retied, or any other intended object or state of affairs. Rather, in the relevant situation, I notice at some pre-reflective and therefore presumably pre-linguistic, and hence pre-semantic level, that my laces need to be retied, so that without another thought I simply act spontaneously by tying them, as though by a reflex action, performing the necessary motions without any evident referential intermediation. I may even be surprised to discover during or after the process that I have been tying or have just tied my shoes. If these examples are convincing, as they should be for anyone with similar experiences, then, although some action-serving intentionality is also meaning-serving, not all meaning-serving intentionality is action-serving. The question is naturally whether in fact the examples do in fact show what they have been interpreted as showing. There is an argument, accordingly, for the alternative interpretation of the relation between meaning-serving and action-serving intentions in option (2). The question is whether it is decisive, or susceptible of being overturned, reversing preliminary intuitive judgments in favor of (2).

It might be objected that the examples do not go deep enough or take into account all crucially relevant nuances of intentionality, action, meaning and reference, to get to the bottom of the question as to how action-serving and meaning-serving intentionality can be understood as related. An argument to be made against (2) in this regard could take one or both parts of the partial overlap of action-serving and meaning-serving intentionality under criticism. We consider these in the same order of exposition as above.

Against (2A): Action need not be physical or involve bodily motion in the ordinary sense, but should be understood generally also to include mental or psychological acts. If I am daydreaming about the Eiffel Tower while remaining virtually immobile on my sofa, then I am certainly engaged in meaning-serving intentionality insofar as my thought is about something, and in particular insofar as I am thinking about the Eiffel Tower. We can safely ignore the fact that idle thoughts are typically accompanied, if not also caused or occasioned by, supervenient upon, or even ontically reducible to, electro-chemical brain events, many, if not all of which, are presumably not under my conscious, deliberate action-serving or meaning-serving intentional control.

Against (2B): We must seriously inquire in this connection whether we can act, physically or mentally, but especially physically, insofar as these categories are justifiably distinguished, without reference, without thinking, and, more pertinently in the immediate context, without any meaning-serving
intention. We encounter most frontally in this regard the difficult and philosophically momentous question as to whether we can ever engage in action-serving intentionality without engaging of necessity as well in meaning-serving intentionality. Can the two be logically or conceptually separated as the argument in support of option (2B) seems to require?

We may try to flesh out the example more fully to further clarify what is entailed. If I tie my shoelaces as an action, then, on the usual analytic expectation, I must intend to do so. Otherwise we usually do not call the motions performed an action. This much is granted, although we also need to ask whether or not such action-serving intentionality can exist or fulfill its role in the explanation of action as an intentional phenomenon without at some level also exhibiting meaning-serving intentionality. If I intend to tie my shoelaces, however rapid or subconscious the intention, implicit or explicit among my acts of consciousness, then my action-serving intention must also ineluctably constitute a specific meaning-serving intention.

I mean to do something specific, at least in the example under consideration. Nor is the description offered as a pun, but in the literal sense of the word. It is to say that my intending thought, whatever its status and circumstances in my cognitive economy, must refer, contrary to the superficial gloss given for (2B), consciously or otherwise, specifically to my shoelaces, my shoes, and the state of affairs in which my untied shoelaces are untied or become tied from a state of being untied. Perhaps I am motivated to tie my shoes by the projection in thought of a plausible but as-yet nonexistent state of affairs which I act in order to prevent, imagining my tripping dangerously over the loose strings to risk injury. I do not need to rehearse the corresponding expressive words and phrases deliberately in thought, in order to have these thoughts, which I reconstruct imaginatively now at my leisure. I do not need to do so in order to refer in fast-moving thought to the intended objects that would be intended in the meaning-serving sense, were I explicitly to run through these terms and sentences tacitly in my mind, as though mentally narrating to myself what I action-serving intend to do in preparing to tie my shoelaces.

**Semantic-Action Intending Relation (3)**

This brings us to alternatives (3) and (4), not precluding the possibility of returning to (1) or (2) if (3) and (4) should turn out to be even more problematic or philosophically unpalatable. Option (3) makes all meaning-serving intentionality a proper subset of action-serving intentionality. A worthy consideration in support of this model is that an extended speech act theory, wherein reference in the symbolic expression of thought in language, and in the making and use of graphic signs, artwork and artifacts most generally, is action or actional. The intentionality of speech acts is distinctively action-serving, with referring and other rightly so-called speech acts being sheltered under the general category as a special kind of action, as something that we do in and with thought and language.
It seems reasonable to suppose that there are referential and extra-referential actions, but that as actions their meaning-serving intentionality is one and all more fundamentally reducibly action-serving. With a pragmatics of meaning built into the concept of meaning-serving intentionality, as a branch of action-serving intentionality, model (3) seems powerfully if not yet decisively preferable at least to alternatives (1) and (2). If referring is something we do, even if only as a mental act or action, and if there are actions that do not require, imply or presuppose any meaning-serving intentionality, then model (3) seems properly suited to account for the meaning-serving intentionality of thought and its expression as an instance, special kind or subcategory, of action-serving intentionality, for one unique (independently) semantically (as others are morally, aesthetically, economically, etc.) evaluated kind of action.

The trouble with proposal (3), despite its virtues, is that it makes meaning-serving intentionality subordinate to action-serving intentionality, with the following implication. If I intend to refer to something as a mental act of intending, effectively intending to do something, even if not performing an intended physical action, the question remains of exactly what it is that I intend to do, of the exact intended object of my action-serving intention. If I think or say that the action-serving intention is to tie my shoelaces, that I tie them, or bring it about that they are tied from a previous state of being untied, then I will have identified the intended object of my action semantically as the meaning-serving intended object of my thought or expression of a decision or action-serving intention to tie my shoelaces. This is precisely what happens whenever I am called upon to articulate or justify rationally what it is I am trying or may have tried to do. It is a question I can intelligibly answer only by applying the meaning-serving intentionality of my reflecting on the intended object of my action, of what it is that I want or wanted to do.

The asymmetry suggests that action-serving intentionality is logically or conceptually dependent on meaning-serving intentionality, and not the other way around. What I intend to do is given by the object of my intending thought when I decide to perform a certain action, even if that action is a speech act or referential effort with a distinctively semantic dimension. My decision is a thought. It is the decision to do something. Its action-serving intended object is referentially determined by the decision’s meaning-serving intending of exactly that object, whether thing or as-yet nonexistent state of affairs. Every action, we are supposing, is intentional in the sense of stemming from a decision to act, where the object of the action is the semantically intended object of the thought. It is in this same way that reasoning and rational discourse have at least an opportunity of linking into the sphere of human action and its consequences, introducing meaning-serving considerations into deliberations about a specific course of action.

Semantic-Action Intending Relation (4)

The preferred model described in the criticism of (3) is the final choice (4). Action-serving intentionality must single out an intended object of action. It must refer in particular to a concurrently nonexistent state of affairs as its
objective. This is the state of affairs that the action is intended to bring about or try to bring about, and which it does bring about if the action succeeds. Actions are not muscle reactions, but intended channelings of energy of one sort or another, mental or perceptibly physical, in an intended direction and toward the accomplishment of a particular intended aim, goal, target or purpose. Actions are the result of decisions to act, or of act-directing thought of some sort, and these thoughts, linguistic or otherwise, involved in actions of all kinds, must also, therefore, depend on meaning-serving intentions. Decisions of whatever form must be about whatever state of affairs is decided upon as the object for an action to try bringing about, and this is unmistakably a matter of meaning-serving intentionality, of the meaning of a decision to act, by which an action-serving intentionality is mobilized.

Nor is the predominance of meaning-serving intentionality over action-serving intentionality subverted by the reflection that referring is also a certain (mental) action or (mental) doing. Assuming that all action is intentional, and that referring is a mental and artistic or symbolic linguistic action, its metaphysical action-serving intended object, once again, is precisely the semantic meaning-serving intended object of the decision to refer, and of the meaning-serving intention to refer to exactly that intended object. What the intended object of an act of referring or other speech act is, can only be answered by the semantic intended object of the referring event, irrespective of the fact that all referring is also an action, occuring as a mental or speech act.

REDUCTIVE MODEL (4) SCHEMA:
ACTION-TO-MEANING-SERVING INTENTIONALITY
Action-Serving Intention
(Decision to Act)
directly Meaning-Serving intends
(Nonexistent) State of Affairs
(Objective) as Direct Intended Object
of the Decision to Act

The relation diagrammed holds when a decision to act action-intends the decision’s meaning-intended object. On the proposed preferred model (4), relating action-serving and meaning-serving intentionality, a decision action-intends only via the mediation of the semantic meaning-serving intentionality of those thoughts in which decisions to act are made. Model (4), as such, best fits the facts even when meaning-serving intentionality is ‘reduced’ to the action-serving intentionality of the undoubtedly mental and sometimes verbal behavioral action of referring in thought and producing its artistic and symbolic expression.
Unified Ontology of Action-Serving and Meaning-Serving Intentionalities

The reduction of action-serving to meaning-serving intentionalities recommends model (4), by virtue of its advantages, over the other exclusive and exhaustive alternatives, including the economical picture it offers of how these types of intending are related to one another. It delivers the explanation despite superficial appearances of conceptual disconnection. Nor should we fail to explain the fact that any action can be understood as a physical expression of thought, as much as its artistic or symbolic linguistic formulation. In deciding to act, the decision arrived at and all the thought process leading to it has a semantic meaning-serving intentionality that locks the resolved-upon action onto a particular intended object, the nonexistent state of affairs that the physical action is intended to realize. Referring and engaging in other speech acts are also actions, doings, without which physical actions cannot be fully understood or explained. However, their action-serving intentionalities are subordinate in every case to a predominant meaning-serving intentionality, so that it is not accidental, purely equivocal, or punning, as Searle advises, to speak in these situations of the meaning-serving intention of what it is we mean to do when we arrive at an appropriate action-serving intention.¹

References


The principle of semantic innocence demands that a particular linguistic expression should have the same semantic value irrespective of the linguistic contexts in which it appears. The importance of this principle lies in the fact that it is also a requirement of a child’s language acquisition capacity. This requirement arises from an adequacy condition posed on a putative theory of language. The condition of explanatory adequacy (see Chomsky 1964 for a more detailed account of this concept) demands that a theory of language must explain how it is possible for virtually all human beings, regardless of intelligence, motivation or even desire, to acquire the native language so regularly and easily.

A theory of language primarily consists of a theory of grammar and a theory of meaning. It is already an established fact by now that a theory of grammar is universal, uniform and single, in other words, it is innocent; otherwise it is not possible for a two-or-three year old child to acquire the complex structure of her native language. So the requirement of explanatory adequacy needs to be carefully addressed in the domain of a theory of meaning. How is it possible for a normal human child to figure out the meanings of as many as thousands of words from the complex noisy linguistic environment to achieve his estimated vocabulary? An attempt to answer this question reflects the need for an innocent semantics.

The issue of an innocent semantics can be addressed by analyzing sub-sentential linguistic items like proper names, definite descriptions, non-denoting expressions, indexicals etc. But I want to restrict the scope of this paper to enquire whether a universal linguistic theory can be discussed in the context of belief, as belief contexts pose a serious threat to the principle of semantic innocence in the sense explained in this paper.

**Frege’s Two-tier Theory of Meaning**

Gottlob Frege, in his theory of meaning, differentiated between the *nominatum* or reference [the designated object] of a sign [name, word combination and expression] and the *sense* [connotation, meaning] of the sign in which is contained the manner and context of presentation. For example, the
nominata of ‘evening star’ and ‘morning star’ are the same, but not their senses. (See Frege 1891 and 1892)

Frege raised several problems that forced him to go beyond a theory of meaning in terms of reference only. His main aim in introducing the notion of sense of an expression was to address the problem of cognitive significance that arises in the context of

1. identity statements,
2. vacuous terms and
3. propositional attitude ascriptions.

For the purpose of this paper I will deal only with Frege’s problem of propositional attitude ascriptions. This problem can be taken as a generalization of Frege’s problem concerning indirect speech to a whole class of expression that is non-truth-functional. It is the puzzle concerning the apparent failure of substitutivity of co-referential singular terms in certain contexts of indirect speech, particularly in propositional attitude ascribing contexts – contexts where the principle of substitutivity of co-referential terms fail.

**The Problem of Substitution**

The principle of substitutivity states that if two terms p and q are extensionally equivalent, then we can replace some or all occurrences of p with q in a statement F(α) without changing its truth-value. Thus,

\[ p = q \Rightarrow F(p) \equiv F(q) \]

Normally this principle does not pose any problem. From the truth of

1. Hesperus = Phosphorus,
2. Hesperus is a planet,

and substitution, we can validly infer

3. Phosphorus is a planet.

However, there seem to be some serious counter examples to this principle. The principle of substitutivity holds only for those propositions which are truth-functional. But for non-truth-functional propositions such as belief sentences the case is completely different. For example, we know that the names ‘Hesperus’ and ‘Phosphorus’ have the same reference, i.e., both of them refer to the planet Venus. Yet, it is not always true that they can replace one another without changing the truth of the proposition. For example, consider the following inference:
(4) Hesperus = Phosphorus.
(5) Riya believes that Hesperus is a planet.
(6) Therefore, Riya believes that Phosphorus is a planet.

If we assume that Riya is unaware of the fact that Hesperus is the same planet as Phosphorus, then (6) becomes false, though (4) and (5) are true.

Therefore, in this context we cannot substitute the name ‘Hesperus’ for ‘Phosphorus’ and arrive at a valid conclusion. One of the most important challenges in philosophy of language is to give a satisfactory account of these contexts where the principle of substitutivity fails (these are known as opaque contexts). If the two co-referential names ‘Hesperus’ and ‘Phosphorus’ make different semantic contributions to the sentences in which they occur, which is a clear violation of the principle of semantic innocence, how will the child understand which theory of meaning to be used to interpret the meanings of those two sentences?

De re and De dicto beliefs

Traditionally speaking, a belief de dicto is belief that a certain dictum (or proposition) is true, whereas belief de re is belief about a particular res (or thing) that it has a certain property. (See Sosa 1970) This distinction can be explained through the following belief-ascriptions:

(7) Riya believes that the earth moves around the sun.
(8) Riya believes, of the earth, that it moves around the sun.

According to the above traditional distinction, the first case is a belief de dicto as it consists in a relation to a certain proposition, and the second case is a belief de re as it consists in a relation to a thing and a certain property.

It can be said that every belief is de dicto in a basic sense, i.e., in the sense that it has a propositional content but it is purely de dicto if and only if an accurate expression of its propositional content is a quantified statement in a sense to be specified below; otherwise it is de re. That is why, there is a special way of reporting a de re belief by saying explicitly that the belief is of an object, such as (8).

According to W. V. Quine (1956, p. 178-180), sentences reporting beliefs and other propositional attitudes are ambiguous. He pointed out the ambiguity in the following sentence:

(9) Ralph believes that someone is a spy.

This could mean either of the following:

(10) Ralph believes that there are spies.
(11) There is someone whom Ralph believes to be a spy.
The first is the *de dicto* sense of believing in which there is a dyadic relation between a believer and a proposition. While in the second case, there is a *de re* sense of believing in which there is triadic relation among a believer, an object, and a property. The distinction between the two can also be seen as a distinction of scope for the existential quantifier. In (10), the existential quantifier is interpreted as having narrow scope, within the scope of ‘believes’:

\[ (12) \text{Ralph believes that } (\exists x) \ (x \text{ is a spy}). \]

In (11), the existential quantifier has wide scope and binds a variable that occurs freely within the scope of ‘believes’:

\[ (13) \ (\exists x) \ (\text{Ralph believes that } x \text{ is a spy}). \]

Accordingly we can say that the *de re*/*de dicto* distinction can be analysed from a syntactic and a semantic point of view. A sentence can be said to be syntactically *de re* if it contains a free variable within the scope of an opacity verb that is bound by a quantifier outside the scope of that verb. Otherwise it is syntactically *de dicto*. So, (12) is syntactically *de dicto*, whereas (13) is syntactically *de re*.

Again a sentence is said to be semantically *de re* if it permits substitution *salva veritate*, which means, if the terms or phrases occurring within the sentence can be freely substituted by co-referential terms without altering the truth-value of the sentence. Otherwise it is said to be semantically *de dicto*. Thus, *de re* belief attributions are referentially transparent whereas *de dicto* belief attributions are referentially opaque.

Quine holds that the logical formulation of the wide scope reading of believing, such as (13), involves quantifying into an opaque context and thus incoherent. He explains it through an example, which is quite similar to the following one.

Suppose, Ralph sees a suspicious-looking man in a brown hat and he concludes that he is a spy. But Ralph is not aware of the fact that the man in the brown hat is the newly elected mayor of the city, Bernard J. Orcutt. Now consider the following two sentences:

\[ (14) \text{Ralph believes that the man in the brown hat is a spy.} \]
\[ (15) \text{Ralph believes that the mayor is a spy.} \]

It is quite clear from the above example that while (14) is true, (15) is false. If it is so, then whether or not Orcutt satisfies the open formula ‘is believed by Ralph to be a spy’ depends on how Orcutt is designated. Orcutt satisfies the formula if designated as ‘the man in the brown hat’, but does not if designated as ‘the mayor’. Quine thus concludes that it does not make any sense to say that Orcutt himself, independent of the way he is being designated, satisfies or does not satisfy the formula; and so, for him, (13) does not make any sense.
If the standard test for distinguishing *de re* from *de dicto* belief attributions is referential transparency or opacity, then we can hold that if Ralph believes *de re* that Orcutt is a spy, then we can characterize Ralph’s attitude by substituting any correct description of Orcutt, like ‘the man in the brown hat’, regardless of whether Ralph would describe Orcutt in that way. The intuition here is that our ascription relates Ralph to the individual in such a way that the particular description or conception that Ralph would use to represent Orcutt plays no role in this sort of ascription. But if Ralph believes *de dicto* that Orcutt is a spy, then the substitution of Orcutt by any co-referential term fails.

According to Fregean semantics, the linguistic formulation of a belief report of the form ‘Ralph believes that x is a spy’ is to relate Ralph to the thought content expressed by ‘x is a spy’ (i.e., the dictum). Since thoughts contain no objects, but representations of those objects, substitution for x by any expression that is merely co-designative with x cannot be expected to preserve the truth-value of the report, as it will change the thought ascribed to Ralph. Thus *de dicto* beliefs are referentially opaque.

Ever since Frege raised the problem of substitution in opaque contexts, philosophers of language and formal semanticists, have been trying to address the problem by specifying the logical form of these contexts. (See Donald Davidson 1968, James Higginbotham 1986 and 1991, Richard Larson and Gabriel Segal 1995, Jon Barwise and John Perry 1983). According to Bertrand Russell (1918, p. 81), what constitutes the puzzle about the nature of belief is that quite often we have false beliefs. He took the example of a false belief like ‘Othello believes that Desdemona loves Cassio’, where the subordinate verb ‘loves’ seems to occur as relating Desdemona to Cassio, but in fact it does not do so. That is why Russell holds, ‘You cannot get in space any occurrence which is logically of the same form as belief’ (Ibid., 82) and by ‘logically of the same form’ he means that ‘one can be obtained from the other by replacing the constituents of the one by the new terms’ (Ibid., 82). Russell thus considers beliefs as new type of propositions–‘new beast for our zoo’–different from every other proposition.

It is obvious that the postulation of ‘new beasts for the zoo’ of facts raises exactly the same problems for explanatory adequacy as with non-innocent theories. If belief-facts are not to be understood with facts that are presented in terms of acquaintance (Russell’s phrase), how does the child know that in interpreting a ‘belief’-sentence he now is confronted with a logically ‘new fact’ for which he needs to marshal wholly new sort of evidence? It seems that appeal to ‘new facts’ simply labels Frege’s problem without addressing it.

Philosophers of language and formal semanticists like Davidson (1968), Higginbotham (1986, 1991), Barwise and Perry (1983) and many others have addressed the issue which has generated a vast literature covering a variety of formal approaches with rich resources; but that is beyond the scope of this paper. All I can say is that these theories either re-state the problem, or address one side of the problem for *de re* beliefs. The issue of whether *de dicto* beliefs can be given satisfactory logical forms at all still remains unanswered.

Therefore, we are left with the choice of accepting either the sceptical solution that there cannot be a theory of meaning for belief sentences (Kripke
1979 and Schiffer 1987, for example) or looking at a new horizon for the analysis of belief sentences so that we can analyse why it is not possible to get a logical structure of belief sentences. This paper chooses the second option because if there is no theory of meaning for belief sentences then we owe an explanation to the fact that we do understand the meanings of those sentences involving beliefs. So something must have gone wrong with our process of analysis. The only justice we can do with the theory of language as well as belief sentences is to change our attitude towards beliefs by treating belief as a ‘local’ concept, in the explained in this work.

**Belief as a Local Concept**

Akeel Bilgrami (1992), following Davidson, draws a close connection between meaning and belief. For him, the literal meaning of a sincere, non-self-deceived, utterance of a sentence by an agent gives the content of the belief that is expressed by that utterance. The central thesis of Bilgrami’s philosophy is a very specific externalist picture of content comprising of two sub-theses – the unity of content and the locality of content.

Bilgrami’s proposed externalism consists of a general notion of the doctrine of externalism – that the contents of an agent’s beliefs are not independent of the world which is external to the agent, i.e., intentionality cannot be completely characterized independent of the external world – added with a constraint that gives a specific reading to the general doctrine of externalism. The constraint is stated as: ‘When fixing an externally determined concept of an agent, one must do so by looking to indexically formulated utterances of the agent which express indexical contents containing that concept and then picking that external determinant for the concept which is in consonance with other contents that have been fixed for the agent’ (Bilgrami 1992, p. 5).

Several things should be made clear in this constraint thesis before going into the explanation of Bilgrami’s theses the unity of content and the locality of content. First of all, Bilgrami points out that ‘concept’ is here used as the counterpart to ‘term’ in the same way as ‘content’ is thought of as the counterpart to ‘sentence’. Secondly, as Bilgrami takes concepts to be externally constituted and contents are composed of concepts, it follows that the contents of an agent’s mind are necessarily public. This amounts to saying that meanings and belief-contents are public phenomena in the sense that they are properties of the agent that is available, in principle, to other agents, who have the capacity for meaning and belief, who can experience the same external environment that constitutes, in general, an agent’s meanings and beliefs.

The publicness of belief and meaning is further explicit from the fact that fixing of an agent’s concepts and contents are done by others. Thirdly, the fixing of externally determined concepts takes place, by looking at the indexical contents in which they occur and the utterances expressing them. Indexical contents and utterances, therefore, provide us with essential clues to an agent’s perceptions and
Bilgrami’s thesis of the *unity of content* claims that there is only one notion of content that is externally determined and that very notion explains actions. It rejects the orthodox distinction between the ‘wide’ and ‘narrow’ content according to which if content is externally determined (‘wide’ content), then it cannot go into the explanation of commonsense psychological behavior, and for that another internal notion of content (‘narrow’ content) is needed. But Bilgrami’s thesis is a *unified notion of content* as it states that though contents are externally determined, it is the same contents that can be used in the commonsense psychological explanation or rationalization of behavior. The thesis of *unity of content* basically tries to put together two widely held assumptions regarding intentional content. The first assumption is that intentional states have a role in the commonsense psychological explanation of action and behavior, and the second assumption is that these contents are externally determined.

While analyzing the reason why philosophers have opposed his thesis of *unity of content*, Bilgrami finds a hidden tension between the above two assumptions. The tension is due to the fact that since the purpose of the explanation of an agent’s action and behavior is to illuminate the behavior of the agent itself, whatever is cited in an explanation must be constituted within the agent and therefore, the examples should show the explanatory failure of externally constituted contents. That is why it is argued that two notions of content are needed. Bilgrami has analyzed the following two examples to show why he is in favor of a unified theory of content rather than a bifurcation of content.

The first example he cited is Saul Kripke’s well-known puzzle about belief. Rather than taking up Kripke’s original puzzle, Bilgrami here works with a variant of Kripke’s example suggested by Brian Loar (1985). In this variant a person named Pierre learns from his nanny, while growing up monolingually in Paris, something which he expresses by saying ‘Londres est jolie’. Later on, he settles in a pretty section of London and after picking up the native language he is disposed to say, ‘London is pretty’. He, however, does not realize that the city he learnt about in Paris is the same city he is living now.

From the close connection between the meaning and belief mentioned at the beginning of this section we can infer that if Pierre is disposed to utter or assent to ‘London is pretty’, then he believes that London is pretty. Kripke formulates this assumption through his *principle of disquotation*. The question that occurs at this point is: does Pierre’s belief have the same content when he assents to ‘Londres est jolie’? According to Bilgrami, those, who will answer this question in the affirmative to show the explanatory failure arising out of externally determined content, will use two further assumptions apart from that of the *principle of disquotation*. One is that externalism is tied to a certain notion of reference, and the other is that translation preserves reference and therefore the truth of the sentence that is assented to. For them, since ‘London’ refers to London and since ‘London’ translates ‘Londres’, the two belief contents must be the same.

Kripke’s original example gives rise to a puzzle by attributing inconsistent beliefs to Pierre. Bilgrami shows that Pierre has the same belief attributed to him earlier in Paris and later in London. The difference is due to the contingent
circumstances where Pierre settles in London. But the source of an element involved in both the examples, which is independent of the contingencies of the examples and which has nothing to do with Pierre being attributed inconsistent beliefs, is that Pierre does not know about the identity that holds between ‘London’ and ‘Londres’. This common element produces a more general difficulty: ‘Belief attributed, under the assumptions mentioned, simply cannot account for the inferences that an agent such as Pierre might make or fail to make … these examples bring out the fact that content attributed under these assumptions will not always be efficacious in explanations’ (Bilgrami 1992, p. 17).

According to Bilgrami, these content attributions are inefficacious due to the insensitivity to Pierre’s different conceptions of London. Pierre conceives of London differently when he is in Paris and when he is in London. And therefore, it is possible for him to fail to know or realize the identity of London and Londres.

Bilgrami’s solution to this problem lies in his proposal of the constraint mentioned earlier which he adds to his theory of externalism. The constraint demands that the external items correlated with an agent’s concepts be described in such a way that it is in consonance with the other contents attributed to the agent and have the effect of including his conceptions of the external items. This constraint makes it sure that we get the right description of the external items to determine an agent’s concepts. For example, if we allow the external elements to enter content under this constraint, the Pierre example will not have any bifurcatory consequences. (For an example of this criticism, see Hanna 2004). In Pierre’s case, the city in question, which is correlated with Pierre’s earlier use of the term ‘Londres’ and later use of the term ‘London’ will be under different beliefs or descriptions of his. In Bilgrami’s notion of externalism, one looks at the agent’s relations with the world around him with the task of imposing characterizations in which the concepts that we find composing the agent’s contents can also be used to describe the world around him with which he relates.

Bilgrami’s thesis of the locality of content is a perfect supplement to his idea of the unity of content. By locality of content, Bilgrami means to say that the explanation of an agent’s actions always takes place at a much more local level than the meaning-theoretic level. The meaning-theoretic level is the level where the theories of meaning do their work. For example, theory of meaning for the term ‘water’ in a particular agent’s language can be seen as an ideally complete specification of his ‘concept’ attributed to him on the basis of all the beliefs, which he associates with ‘water’. But the entire aggregate of beliefs is not required for the attribution of specific content to explain behavior and here comes the importance of the locality of content. At the local level one sifts out of the pool of resources consisting of various beliefs only those which are relevant in order to attribute specific contents for the explanation of action.

Thus, it is very important to understand that a distinction needs to be made between two levels – the level of a theory of meaning, on the one hand, where all the beliefs are aggregated in the specification of a given concept; and the level of content, on the other hand, where actions are explained. In the explanation of an
action, belief-contents are invoked, and these belief-contents are composed of concepts, but *not of all the concepts specified at the meaning-theoretic level, rather only of those concepts selected at the local level.*

For our purposes, Bilgrami’s handling of the problem of belief attribution leads to several crucial consequences.

A. Bilgrami’s *position* with respect to content of beliefs, namely, unity of content, is a species of semantic innocence advocated throughout this work. Bilgrami is opposed to the distinction between narrow and wide contents, which, in semantic terms, translates into a distinction between Fregean (narrow) and Russellian (wide) semantics (see Fodor 1994). Bilgrami not only rejects the distinction, he holds that the content of belief is basically ‘externalistic’ in character. To that extent he is not only advocating semantic innocence, he is opting for referentialism as the only candidate for semantic innocence.

B. Bilgrami advocates a central and novel distinction between ‘meaning-theoretic’ explanation and explanation of content, especially for contents of beliefs. This is where Bilgrami’s suggestions take a decisive *sceptical turn* because, in effect, he is arguing that the contents of beliefs escape meaning-theoretic explanation. In so far as theories of logical form are viewed as meaning-theoretic explanations, contents of beliefs cannot be captured in terms of the *form* of belief-propositions. Translated in our terms, semantic innocence in belief-contexts is not compatible with their logical form.

**Concluding Remarks**

What then is the locality of content in belief-ascriptions? According to Bilgrami, ‘the entire aggregate of beliefs is not required for the attribution of specific content to explain behavior and here comes the importance of the locality of content’. Bilgrami’s solution requires that something like the notion of ‘what Pierre has in mind at t’ determines the content of Pierre’s beliefs at t, where ‘has in mind at t’ is to be understood locally. That is, the primitive notion for determining the content of Pierre’s beliefs is *not the semantic value of one of its constituents* (for example ‘London’), however constrained at the circumstance of evaluation, but that content of ‘London’ that Pierre has at t. Once the effect of locality is understood in this way, the account of Pierre’s beliefs becomes a ‘psychological account’ in Schiffer’s sense, although we can continue to assume that each of these states of Pierre are driven by externalistic content. In other words, there is no necessary shift to ‘narrow content’ just because the notion of what Pierre has in mind has been taken to be the primitive. However, as Schiffer’s incisive analysis revealed, any psychological account of the sort just described is incompatible with the demands of a (general) theory of meaning that is supposed to hold for a *language*, not just for Pierre’s current beliefs. Therefore, given the link between the locality conditions governing Pierre’s psychological states, and
the limits of formal enquiry on the content of those states, we can see why a
theory of beliefs does not fall under putative theories of meaning for language that
we can currently make sense of.

Generalizing to Frege’s example, we can now say that in evaluating S’s belief
that Hesperus is a morning star, we need not simultaneously take into account S’s
other belief that Phosphorus is an evening star. In each case we need to treat S’s
use of ‘Hesperus’ and ‘Phosphorus’ separately in terms of the locality conditions
governing their use, even if in meaning-theoretic terms Hesperus = Phosphorus.
Therefore, S cannot be viewed as saying, of the same star, that it is at once a
morning and an evening star; the locality conditions governing S’s use of these
names does not guarantee that one can be substituted for the other. By parity of
reason, it cannot also be ruled out that S might be so governed by locality
conditions that he holds that Hesperus = Phosphorus (just as we do!). In that case
S is not likely to say of the same star that it is at once a morning and an evening
star. If, despite these precautions, S does in fact say so, he can legitimately be
charged for holding (apparently) inconsistent beliefs. But these decisions depend
tightly on S’s mental biography, rather than on the meanings of terms.

We could say that we might have entertained a philosophical view that
contents of beliefs also fall under general theories of language in terms of
valuation of the predicate ‘believes that’ because giving a theory of meaning for
an expression X is the only way of explaining what we understand on
saying/hearing X. The fact that the explanatory scope of our favourite theories of
meaning – referentialism, in this case – can be extended to cover embedded ‘that’-
clauses – such as ‘sees that’ and ‘says that’ – might have encouraged us to extend
the scope further to ‘believes that’ as well. After all, if language is used to talk
about the world, then the general conditions governing that talk (=theories of
meaning) ought to cover every piece of talk including those involving ‘believes
that’; how else do we make sense of that talk?

The locality of content idea articulated above can be used against the
generality of content implicit in the preferred philosophical concept of belief. If
this approach is plausible, then the problem with *de dicto* belief proposition and
the substitution problem thereof could be traced to the generality assumption
implicit in the philosophical concept of belief. We can thus say that the problem
may not arise if we give up the generality requirement and with it the meaning-
theoretic requirement – that is the requirement of logical form – imposed on belief
*de dicto*.

The prevalence of the locality condition in determining the content of belief
reports opens new directions for philosophical clarification of the concept of
belief. Although beliefs are reported in language, the content of belief do not seem
to obey the classical conception of semantics as language-world connections,
where the ‘world’ is typically understood as the external, mind-independent world
– referentialism. The locality condition suggests that what rules in belief contexts
is the ‘constructed world’ of the agent of beliefs – the ‘thought’ world. This was
Frege’s central insight. But the (mind-internal) world that essentially determines
the content of beliefs does not pervade the entire length and breadth of language –
the locality; hence, it defies the classical conception of semantics. It could be that
the ‘world’ of beliefs is a different kind of world; it is my world, as Wittgenstein would say. A study of this insight is beyond the scope of the present project.

References

Peirce on Continuity: Vindication of Universals against Nominalism

Paniel Reyes-Cárdenas, The University of Sheffield, UK

In his recent book “Peirce and the Threat of Nominalism”, Paul Forster (2011) argued how Peirce understood the nominalism. Forster also explained that the nominalist move of individualising concepts for collections could cost denial of properties of true continua. Peirce showed some vibrant issues both in mathematics and in metaphysics, as for example, the classic one of universals. That work, however, is still incomplete; needs to be clarified how these ideas bind with Peirce’s doctrine of ‘scholastic realism’. Peirce’s own theory of continuity might be the answer: I trace the origins of this doctrine in the theory of multitude. I show how the theory of continuity frees analysis from constraints of nominalist theories of reality. Peirce’s theory of multitude, instead, can be derived with mathematics: By drawing in the work of the ways of abstraction in diagrammatic reasoning made by Sun Jo Shin (2010) and in continuum theories by Cathy Legg (2010) I show the conceptual device of “diagrammatic reasoning” as a plausible pragmatic tool to represent continua and make sense of Peirce’s metaphysics. The analysis of continuity is a perfect example of how the method of diagrammatic reasoning helps to unblock the road of philosophical inquiry as well as clarify other problems as for example, the applicability of Mathematics. I conclude explaining that general concepts define ‘continua’. The properties of true continua are not reducible to individuals they comprise. Continua are the universals of metaphysics; they are intelligible and necessary to ground the science of inquiry.

Keywords: Universals, continuity, scholastic realism, applicability of mathematics, diagrammatic reasoning, Peirce.

Introduction

There is probably a huge task in undergoing a unified account for contemporary philosophical methodology. The traditional “analytic-continental” distinction never really was an accurate account of that state of affairs. Nonetheless I would like to put forth some ideas that are present in the
contemporary philosophical atmosphere: the wish for a 'pragmatic turn' in philosophy, as Richard Bernstein (2010) had written in a rather still recent book. The turn is about make the values of both philosophical traditions (analytical and continental) converge into a sensible unified account of inquiry. Peirce, as Karl Otto Apel recognised several years ago now, is a philosophical figure of particular interest to understand the needed turn, he considered that philosophy needs to get rid of any possible block to the way of inquiry, and created a method for philosophy called ‘pragmatism’. The method gravitates around the ‘pragmatic maxim’ that works as its core. The pragmatic approach to philosophy concentrates the maxim within a ‘road of inquiry’ and acts as a normative science that prescribes principles for inquiry. This was recognised, in Peirce’s time, as a legitimate sense of the term ‘Logic’. Thus, understanding Logic as a ‘science of inquiry’ prescribes a correct account of methodology for any object of human cognition, starting off from the aim of achieving a ‘scientific metaphysics’. In this article I will describe features of the pragmatic maxim unwrapping its heuristic value against its nominalist rivals and finally showing how the grounding of the science of inquiry from a pragmatistic point of view needs mathematical diagrammatic reasoning.

The realism of universals is a common topic in the pragmatist scholarship of Peirce. This realism is often misunderstood as a platonistic untenable view. My aim in this paper is to show how Peirce’s plea against nominalism is a feasible account for aiming a scientific metaphysics grounded in mathematical diagrammatic reasoning and requiring an account of continuity and universals.

**The Pragmatic Maxim**

The core of the pragmatic approach to philosophy initiated by Charles Sanders Peirce (1839-1914) is the 'pragmatic maxim'. This was officially born in a celebrated essay called “How to make our ideas clear” (1878). The formulation of the maxim goes like this:

“Consider what effects, which might conceivably have practical bearing, we conceive the object to have. Then, our conception of these effects is the whole of our conception of the object.” (CP 5.402, 1878)

The slogan mentions “conceivable practical bearings”; the sense of the expression is to urge not only for empirical immediate consequences. The maxim suggests that those bearings are to be understood as having implications for what we should do or expect. This is why years later Peirce felt necessary to clarify the maxim in the context of the theory of signs:

“The entire intellectual purport of any symbol consists in the total of all general modes of rational conduct which, conditionally upon all
the possible different circumstances and desires, would ensue upon the acceptance of the symbol” (EP 2: 346)

The pragmatic maxim, as a critical tool, is drawn to overcome any unjustified dichotomy in knowing reality. For example, dichotomies can be found between impressions and things-in-themselves. The maxim, therefore, applies to a theory of inquiry as well as the foundations of metaphysics. Now, a theory of reality must acknowledge the conditions that make inquiry itself possible. Consequently, should metaphysical inquiry be possible, foundation and basis for metaphysics that is not viciously circular must be given. The experimental method implicitly affirms the reality of generals or universals (see CP 5.494). That is part of Peirce’s concern to use the maxim properly and render realism plausible. The pragmatic maxim acts as a norm for inquiry that regulates theories about methods. The aim of systematising those norms by using the maxim will be revisited in the next section.

Metaphysics and the Science of Inquiry

For Peirce, Logic is the science that deals with the principles of right reasoning (W3:244). Logic, so conceived, involves principles of formulating hypothesis, deducing testable consequences from them, and evaluating their truth or falsehood. The principles mentioned are of the following kinds: abduction, deduction and induction; these are kinds of inferences. Things being so, inferences altogether account for all ways of growth in cognitive lives. Historically speaking, there is a wide tradition of empiricism that recommends nominalism as the metaphysical stance to take. Nominalism generates a kind of metaphysics that focuses on the a priori conditions of knowledge. The acceptance of nominalist conditions as principles, however, is problematic: the principles are supposed in the justification to accept them. For Peirce that constitutes what he despised as “ontological metaphysics”. Peirce thinks that:

“The only rational way [to do metaphysics] would be to settle the first principles of reasoning and that done, to base one’s metaphysics on those principles” (CP 2.166, 1902)

The principles have not to be necessarily a priori. Principles emerge from the needs for explanation that inquiry prompts in us: Peirce thinks that Metaphysics is a puny and rickety science because it has been coined out of nominalistic scruples. If we make it out as a science based on experimentation and observation instead, metaphysics is a science that finds its principles in experience broadly construed.
Nominalism and Inquiry

Nominalism is a doctrine that affirms reality comprises only individuals; it also denies that there are laws operating in reality. The nominalist believes that a complete account of reality can be formulated only by enumerating individuals and their traits. This implies no laws, general concepts or abstract objects identified as real.

The nominalist rejects the reality of abstract objects and universals. One channel of her rejection goes along by only giving the status of ‘real’ to individual objects within the range of sense data. This nominalism is a doctrine otherwise known as empiricism. When methodological philosophy, a science of the principles of inquiry, is formulated in these terms, it is equivalent to a further natural science amongst other natural sciences. These are the characteristics for which, nonetheless, a nominalist consideration of metaphysics as natural science is problematic if proposed as science of inquiry (Cf. Forster 2010):

i. Nominalism implies that principles of inquiry are justified a posteriori: it is not clear, though, why these should be salient amongst other principles of other sciences that are a posteriori too.

ii. The principles of inquiry, to be applicable as principles need to be formal whereas, in the natural sciences, are material.

iii. If the principles are discovered they can be interpreted culturally specific rather than universal.

iv. The view implies that the principles are contingent. However, they should represent our best account of necessity: as opposed to contingent they have to be necessary.

v. The nominalist takes the principles as descriptive, but they need to be normative if we need a way to attain truth reliably, i.e., by a prescriptive way.

These problems make the justification of principles based in natural science viciously circular: they pretend to give foundation to metaphysics, but fail. Nominalism as natural science is restricted to a particular, concrete, and non-general sense of inquiry.

Another way to provide a nominalist stance is by developing a theory of inquiry as ‘a priori science of the mind’. This perspective makes the ‘science of inquiry’ a priori, formal, universal and necessary. The problem in this nominalism, though, stems from the grounding of the principles on ‘intuition’. Intuition renders the science of inquiry subjective. If the science of inquiry is subjective then we are unable to distinguish, for example, a self-evident principle from another that appear to be so. Peirce diagnosed this strategy as ‘method of tenacity’ (W2: 212). This nominalism does not allow for independent, inter-subjective experimental testing of the principles of inquiry.
Peirce’s Proposal: Synechism, the Theory of Continuity

Peirce believed that ontological/a priori metaphysics blocks the way of inquiry. Nominalism in all its forms gets to a halt by postulating incognizables. Scientific metaphysics, on the contrary, struggles to achieve the general features that are present in experience broadly construed: by and large. Peirce’s proposal for a scientific metaphysics uses categories as descriptions of the universes of experience. His particular list of categories renders the hypothesis of reality plausible and yet does not commit to the problems of nominalism. Peirce’s proposal is a posteriori. This implies that fallibilism applies to scientific metaphysics. The fallibilism of the theory does not lead to scepticism but to context-sensitive inquiry. Consequently, the principles of this new metaphysics are no more than fallible hypotheses and not metaphysical necessary axioms. Peirce considered three hypotheses as equivalent to the first principles. These three hypotheses are doctrines named as:

1. Synechism (or the theory of continuity)
2. Tychism (or the theory of real chance)
3. Evolution (or the theory of the tendency to habits)

Let me analyse the first one. By doing so, it will become apparent that they are not exactly equivalent to traditional realism. They might be recognized under the same spirit of traditional accounts of realism, though. These doctrines are all but absolute: they cannot be absolute but relative; part of inquiry conceived as experimental and a posteriori. Peirce’s cosmology is inquiry through these three doctrines, principles discovered by observation and experimentation. I shall explain how his metaphysics is comprised by these doctrines. First and foremost, the doctrine of continuity, named by him “Synechism” is the more prominent and important of this metaphysics. In Peirce’s system Synechism is the most important hypotheses of metaphysics:

“Synechism is not an ultimate and absolute metaphysical doctrine; it is a regulative principle of logic, prescribing what sort of hypothesis is fit to be entertained and examined” (CP 6.173)

Peirce thought that modern philosophers blundered in their adoption of the paradigm of geometry. He wanted to draw attention to the kind of inquiry carried by Aristotle and the medieval philosophers, especially focusing in Duns Scotus’ account of metaphysics. Scotus’ account included, amongst other things, theories of individuation and an account of universals that reconciles the first principles of Aristotle’s metaphysics taking them not as objects of a priori knowledge. In this regard Peirce was very loyal to Scotus’ approach.

Whereas the other principles of his metaphysical system are derived from evolutionary biology, the case of Synechism is peculiarly and specifically derived from the study of mathematics. Peirce developed his Synechism by studying the mathematical concept of ‘continuum’. The best account of
Peirce’s concept of continuity and the *continuum* is contained in the Cambridge Lectures of 1898. To illustrate, Peirce analyses the structure of a line. Lines are usually considered as collections of points. Peirce resisted that conception of a line:

“[N]o point in this line has any distinct identity absolutely discriminated from every other” (RLT 159)

The problem of conceiving a line as a collection of actual points lies in this: the discrimination of one individual point actually separates that point from the other points. This leads to the kind of paradoxes such as the Achilles’. The problem of the concept of a line as a collection is this: it is unable to distinguish a collection of actual points from a collection of potential points. The mistake is due to this: discontinuity is created by actual points. This draws attention to the main feature of the line: its continuity. A line, thus, can be better conceived as a continuous of *infinite* number of potential points. Peirce defines the mathematical continuum as a blend where potential points are “welded”:

“Namely, a continuum is a collection of so vast a multitude that in the whole universe of possibility there is no room for them to retain their distinct identities; but they become welded into one another. Thus the continuum is all that is possible, in whatever dimension it be continuous. But the general or universal of ordinary logic also comprises whatever of a certain description is possible. And thus the *continuum* is that which the Logic or Relatives shows the *true* universal to be. I say the *true* universal; for no realist is so foolish as to maintain that universal is a fiction.” (RLT 160)

Allow me to explain Peirce’s idea of the continuous line by means of picturing a line in which a cut is carried out: Putnam (1995, 13) and others identify this as a “Dedekind Cut”. The cut divides a line into two segments. Let L be the left segment and R the right segment. The cut is applied to divisions where a line tat can be measured in terms of rational numbers has the four following properties:

1. L and R are not empty;
2. If a number belongs to L, then so does every smaller number;
3. If a number belongs to R, then so does every bigger number;
4. Every number belongs to exactly one of the sections.

The line presents us with a seemingly paradoxical aspect: should we recognise it as measurable (in terms of numbers and points) then the line has points in itself; but if we carry out the Dedekind cut what results is two lines in
each segment instead of a single line. An Aristotelian conception will help Peirce to overcome this problem:

“In like manner, the potential aggregate of all the abnumerable multitudes is more multitudinous than any multitude. This potential aggregate cannot be a multitude of distinct individuals any more than the aggregate of all the whole numbers can be completely counted. But it is a distinct general conception for all that… a conception of potentiality” (RLT 160)

Peirce uses the concept of ‘potentiality’ to discriminate the aspect for which the line is potentially divisible infinitesimally in different segments. This possible real quality does not mean an actual division, but a potential realization. When the Dedekind cut is carried out what results in actual lines is the development of discontinuity. Discontinuity, as a concept, is understood in terms of the previous continuity. The division, therefore, makes sense because is carried out in the framework of the still not divided and potentially divisible.

Peirce emphasises that the true universal/real general (the metaphysical concept) is of the nature of this continuum. The universal/general is to be understood as a continuum of potentials in some dimension, the continuum gets actualized when we discriminate one of those potentials. When we distinguish a case of a type what we do is something analogue to the ‘Dedekind cut’ of a line: we discriminate an individual that can only be understood upon the background of a continuous essential property.

Peirce’s Synechism has an important asset as a theory of universals: ‘continua’ are concepts that allow vagueness. Vagueness is a very relevant aspect of an account of realism: prevents from falling into “absolute conceptions of reality”. Absolute conceptions of reality believe that every answer has a concrete answer: either true or false. Absolute conceptions predicate bivalence. There are highly problematic cases for absolute theories. Consider the question that asks whether Caesar sneezed two or three times the morning (afternoon?) that crossed the Rubicon. No matter how well and long we inquire, these facts are lost to us. An account of vagueness, instead, can give us the elements necessary to account for reality even in its dynamical and changing aspects. Conceiving universals as ‘continua’ renders us possible to inquire about something gravitating towards a true that is not a single fact, but an ‘ideal’ limit. Vagueness does not mean obscurity. It is actually quite the opposite, for Peirce, Scientific Metaphysics, through Synechism can account for the pervasive different aspects of universals in our inquiries into reality. Peirce went further here, he not only offered an interesting approach to metaphysics, but offered an account in which metaphysics and mathematics converge in their foundations: through the study of the mathematical continuum we can also account for the continuum in reality, and, finally, this study starts off as a study of Diagrammatic Reasoning, which is a study of how mathematical continua can be inquired through experimental means, even being completely abstract From an epistemological angle, the study of
mathematical continua is approached with diagrams, Peirce’s particular paradigm of epistemology in mathematics. Peirce cared to account for that in what he called as “Diagrammatic Reasoning”.

**Diagrammatic Reasoning**

Peirce’s option in mathematical epistemology is to propose mathematical diagrammatic reasoning. When reasoning with diagrams we do not render the principles of inquiry viciously circular. We do not rely on intuitions either:

“What is needed above all, for metaphysics, is thorough and mature thinking; and the particular requisite for success in the critic of arguments is exact and diagrammatic thinking’ (CP 3.406)

In order to make sense of how there can be a process of reasoning wide enough to include all the desiderata we enlisted before let us describe the process of mathematical diagrammatic reasoning (following Hoffman 2003, 121-143):

1. Constructing a diagram by means of a given representational system (Euclidean geometry, Peano or Peirce Arithmetic, a language, some computer software... etc). The construction is motivated by the need to represent relations.
2. Experiment with those diagrams, as long as they define constrains that determine the outcome of experiments and then impinging something that in the actual world will be an inevitable experience.
3. Observing the results of experimenting, gathering a new insight on getting something out of the outcome of diagrammatisation.

**Continuity and the Problem of Universals**

Peirce’s answer to the problem of mathematical inquiries is closely linked to his belief on true continuai. Peirce denied that a continuum in mathematics is a collection of individuals, which is the point of view of the nominalistic approach. He rather proposed conceiving continuum as series in which the members are specified after recognising the continuity. So, for example, in the natural number series the properties of the natural numbers are coming out from the series and not from the particular numbers attached together, otherwise known as discrete quantities. Peirce thinks that this kind of mislead path conduces to the sort of paradoxes of Achilles and the Tortoise. Peirce thinks that mathematical induction, that he called ‘Fermatian inference’
(Peirce, NEM 3:49, 1895), is valid for any collection whose members in one to one correspondence with the natural numbers.

“For Peirce, then, the example of Fermatian inference shows that the nominalist is wrong to think that the only valid principles of reasoning about infinity are those that apply to finite collections.” (Forster 2010, 47)

Forster speaks about an epistemological aspect of the continuity: should we understand what the notion is about we firstly have to explore the periods of change in Peirce’s ideas about continuity in mathematics (see Potter 1996, 117s)

The nominalist definition of continuity squares generals as collections of individuals, and this in turn led to the conclusion that series comprise individuals in collections. This turns out to be extremely problematic not only for mathematical series that are divisible virtually to infinity, but also for true continua experienced in the process of scientific inquiry, cases where series of common events obey to a law that seems to act really and fundamentally. Peirce thinks that the mathematical analysis of continuity is a perfect example of diagrammatic reasoning unblocking the road of inquiry and grounding the way for a science of that inquiry pragmatically. The nominalist blocks the way of inquiry in this sense: as defining generals as collections accounts for the properties of the generals in terms of sums of the properties of the individuals, but they literally emerge as different, especially when it comes down to sciences like physics, where the generality is not a function of the individuals ingredient in the general.

Now the way to come across this kind of fundamental aspect of the continua is by means of diagrammatic reasoning. Reasoning can be defined in terms of inferences, but carrying out inferences means realise operations that often involve more than one single medium, Peirce thinks that the kind of inquiry that metaphysics requires. For Peirce, as we said, Logic is the science that deals with the principles of right reasoning (W3: 244), principles of formulating hypothesis, deducing testable consequences from them, and evaluating their truth or falsehood. This sense of logic is obviously different to the contemporary use; it is an account for the kind of inferences that we can carry out by deduction, induction and abduction. Diagrammatic representations allow us to carry that kind of reasoning in mathematical expressions, not only deducing, but also formulating hypothesis and inducing mathematically. Mathematics holds a normative character for anything that aims to be a ‘science of inquiry’:

“Although mathematics deals with ideas and not with the world of sensible experience, its discoveries are not arbitrary dreams but something to which our minds are forced and which were unforeseen” (Peirce 1894; N2.346)
Peirce denied that mathematics depends in any form of intuition of space and time. Peirce coped with the problem of analytical of synthetic propositions within the mathematical language. The problem of nominalism is that although it could make the ‘science of inquiry’ a priori, formal, universal and necessary, its use of intuition renders the science of inquiry subjective. We do not have means to distinguish a self-evident principle from one that seems so, thus, it does not allow for independent, inter-subjective and experimental testing of the inferences. Peirce harshly criticised nominalism based on intuition as a kind of method of tenacity (W2: 212) and as an ungrounded Cartesian assumption that we can have “cognition without signs (EP 11-13). Peirce thought, instead, that mathematics is not a science of facts but of hypotheses and abstractions. Concerning the truth of mathematical statements, Peirce thought that mathematical necessities are somehow previous to truth, this opinion develops in the desire for bring mathematical exactitude into philosophy.

The Subject Matter of the Science of Inquiry with a Foundation on Synechism

These are the ways in which the diagrammatic approach meets the demands summarised before:

i. If under conditions of diagrammatic reasoning we establish principles that would count as a prescriptive aspect of a theory of inquiry, even if our inquiries were experiential, observational and experimental.

ii. A diagrammatic approach delimits possible states of affairs rather than determinate what is the case. Consider the case of a theorem, for example, Pythagoras’: derives laws that apply to every possible triangle, even if the theorem does not speak about any triangle in particular.

iii. The principles of diagrammatic reasoning hold in all cases: rational inquiry as diagrammatic is unlimited and universal in scope. This is a desideratum required for an account of the science of inquiry.

iv. The principles derived from diagrammatic reasoning are necessary: they hold for any possible world in which we are interested in finding truths throughout rational inquiries.

v. The science of inquiry should be normative. The principles of diagrammatic reasoning are rational proofs that become generally prescriptive. These are the case whenever a universal is to be represented rationally as a continuum.

Now, after these premises, the account based on diagrammatic reasoning (1) preserves the characteristics we need for the science of inquiry, (2) avoids the circularity of justifying principles of Inquiry by appeal to claims of the
same principles, and with especial relevance (3) allows the sort of independent, repeatable, inter-subjective testing that is the hallmark of rational inquiry. Finally, let me offer a long quote that includes examples in which experiments and diagrammatic reasoning converge:

“To say that a quadratic equation that has no real roots has two different imaginary roots does not sound as if it could have any relation to experience. Yet is strictly expectative. It states that would be expectable if we had to deal with quantities expressing the relations between objects, related to one another like the points of the plane of imaginary quantity. So a belief about the incommensurability of the diagonal relates to what is expectable for a person dealing with fractions; although it means nothing at all in regard to what could be expected in physical measurement... Riemann declared that infinity has nothing to do with the absence of a limit but relates solely to measure. This means that if a bounded surface be measured in a suitable way it will be found infinite, and then if an unbounded surface be measured in s suitable way, it will be found finite. It relates to what is expectable for a person dealing with different systems of measurement” (CP 5.541)

The ‘science of inquiry’ is an application of the pragmatic method: works in terms of conceivable practical bearings; these are not reduced to empirical experience. Continua have a prescriptive aspect: not grounded in intuition and still good enough to be formal principles:

“There is no Kantian noumenon. If this is so, the Real constitutes a network of relations such that everything is connected with everything else or, to put it another way, the Real is everywhere continuous... this continuous Real is systematically explored through abduction, deduction, and induction.” (Potter 1977, 75)

Conclusion

Thus far, I reviewed Peirce’s proposal for metaphysics. His unique proposal is an interestingly intricate response to the foundations of metaphysics and the relationships of metaphysics with mathematical thought; all of these in an account of universals. Two particular conclusions seem relevant here in order to capitulate what has been said:

• A Metaphysical answer: The all-pervasive free-standing structures of mathematics are primarily true continua. These same structures are the universals we find in scientific inquiry.
• An Epistemological answer: we can access to continua with self-controlled inquiries by means of diagrammatic reasoning
References


(1894) As a Manuscript in: Richard Robin (1967) Annotated Catalogue of the Papers of Charles Sanders Peirce. Boston: The University of Massachusetts Press [N2; Ms 346]


Part C: Philosophy of Science and Philosophy of Mind
Rosenthal argues that it is not possible to think about a particular mental state token without thinking about the subject whose mental state it is. I refer to this as the token condition. In this paper my purpose is to undermine the token condition. I discuss two reasons for the token condition. The first of these reasons consists in the what-it-is-likeness of experiences and the second consists in the strong conceptual relation between experiences and their subjects. I attempt to undermine this relation and suggest an account of the what-it-is-likeness of experiences without relying on the subject’s point of view in order to undermine the token condition. I then suggest that a slightly different version of Rosenthal’s higher-order thought theory of consciousness is compatible with the possibility of thinking about one’s particular mental states without thinking about oneself. Finally I argue that even if the token condition is true, there is an asymmetry between the way I think about my mental states and the way I think about another person’s mental states and given this asymmetry, the token condition cannot be said to apply to all mental state tokens one thinks about.

Keywords: Mental states, higher-order thought, subjects

Introduction

Rosenthal (1997, p.741) says that it is not possible to think about a particular mental-state token, as opposed to thinking simply about a type of mental state, without thinking about the subject whose mental state it is. In other words, when I am in pain and think about this particular pain state, I must think about it, the token state, as my pain state. So, I need to think about myself although not necessarily in an attentive manner.

However what it is that one is thinking about when one is thinking about oneself is far from clear and yet rarely challenged possibly due to the strong influence of Descartes’s Cogito Ergo Sum. At the same time though Cartesian Egos have long been the unwanted guests in our ontologies. And despite Descartes’s influence, one does see in Hume’s (1739-40/1978, p.252) famous words a significant dissatisfaction with the taken for granted experience of oneself; ‘For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or
shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe anything but the perception.’

Undoubtedly the existence of a self can neither be confirmed nor refuted by there being an experience of it or lack thereof. Nevertheless, the controversy over having such an experience, not to mention the controversy over both the nature of the experience granted we have it, and the nature of the self granted there is such a thing, does give us a good reason to see if we can explain the mental phenomenon of thinking about one’s mental states without thinking about oneself.

In what follows, after briefly giving some historical background, I discuss two reasons for the view that it is impossible to think about a particular mental state token without thinking about the subject whose mental state it is. Henceforth I shall refer to this view as the token condition. The first of these reasons consists, in Nagel’s (1974) famous words, in the what-it-is-likeness of experiences and the second consists in the strong conceptual relation between experiences and their subjects. I attempt to undermine this relation and I suggest an account of the what-it-is-likeness of experiences without relying on the subject’s point of view in order to undermine the token condition. Although Rosenthal (1997, p.741) embraces the token condition, I suggest that a slightly different version of his higher-order thought theory of consciousness is compatible with the possibility of thinking about one’s mental states without thinking about oneself. Finally I argue that even if the token condition is true, there is an asymmetry between the way I think about my mental states and the way I think about another person’s mental states and given this asymmetry, the token condition does not apply to both.

**Historical Background**

It is possible to trace the roots of the token condition as far back as to Aristotle. Aristotle (*Sense and Sensibilia* 448a26-8) suggests albeit in an if-clause that it is impossible for a person to perceive something and be unaware of his existence. In *Nicomachean Ethics* too, Aristotle (1170a31-3) says that if we perceive, we perceive that we perceive and if we think, we think that we think and to perceive that one perceives or thinks is to perceive that one exists. The idea can also be found in Locke (1694/1975, p.115) when he says ‘thinking consists in being conscious that one thinks.’ Within the phenomenological literature, Heidegger (1982, p.132) says that in having a feeling for something, there is always a self-feeling too. Gallagher and Zahavi (2008, p.46-7) cite this and various other remarks in the phenomenological literature approvingly and describe awareness of oneself as ‘an intrinsic feature of primary experience.’ In addition Zahavi (2006) argues that consciousness essentially involves self-consciousness.

There are similar remarks in the analytical tradition too. Goldman (1970, p.96) says that the process of thinking about something carries with it a non-
Thinking about Mental States

reflective self-awareness. Flanagan (1992, p.194) speaks of a ‘low-level self-consciousness involved in experiencing my experiences as mine.’ And Kriegel (2003, 2004) argues that all forms of consciousness depend upon a peripheral awareness of oneself as the subject of one’s mental state. So there is some consensus with regard to the general idea behind the token condition.

Reasons for the Token Condition

In considering the reasons for endorsing the token condition, the first question to ask is why thinking about a particular mental state token implies thinking about its subject. One obvious response would be that in thinking about a particular mental state token, one also thinks about what it would be like for someone, viz. for its subject, to have it. In order to think about a particular mental state token of say the olfactory sensation of recently mowed grass, one needs to think of the olfactory sensation as somebody’s sensation with its unique phenomenology since that’s what makes it a particular mental state token and not a type. Hence the what-it-is-likeness of conscious states seems to be one of the reasons for the token condition.

Another reason for the token condition consists in the strong conceptual link between experiences and subjects of experiences. The concept of an experience entails the concept of something having the experience. In other words, where there is an experience, there is a subject of experience. After Strawson (2003, p.280) I call this the subject thesis. Our understanding of beliefs, desires, memories, emotions, perceptions etc., ordinarily assumes something that believes, desires, remembers, feels, perceives etc. even if there seems to be no consensus regarding what the nature of this thing that desires and remembers is. Typically it is either brains or Cartesian egos but the problems with these suggestions are well known. (Crane, 2003) Nevertheless, James (1890/1950) argues that the elementary psychic fact is not just this or that thought, but someone’s thought, every thought being owned (p.226). Shoemaker (1986) argues that experiencing involves its subjects ‘as intimately as a branch-bending involves a branch’ (p.107). The notion of ownerless experiences seems counterintuitive.

Also when it is taken into consideration that mental states are states, one could argue that a state is a state of something – in this case the subject – and that it doesn’t have an existence outside of the thing it is a state of. However, while this is true for all states including mental states, it does not immediately translate into the subject thesis. For instance a roll of thunder, insofar as it is a state of the weather, does not have an existence outside of the relevant medium of atmosphere. But we do not thereby posit a subject of the thunder, or we do not necessarily think of something that thunders.

However, one might argue that mental states are supposed to be different from states of the weather and that we do not think of mental states in the way we think about weather states. Strawson (2003) argues, if there is pain, but nothing that suffers from it, the incentive to seek relief from the pain will be
meaningless. (p. 281) But perhaps it is part of a pain-state that it calls for a relief, in which case the need to ease a pain would not necessarily be dependent on there being something suffering from the pain, but on the state itself. Still, it seems only natural to endorse the immediate inference from mental states to subjects of those mental states, and the inference is rarely second-guessed. Hence I take the subject thesis as the second reason for the token condition.

**Undermining the Token Condition**

Despite the grip of the subject thesis, there are some who are skeptical about the inference from mental states to subjects of those mental states. For instance, Nietzsche (1968, p.268) says ‘But that means positing as “true a priori” our belief in the concept of a substance – that when there is thought there has to be something “that thinks” is simply a formulation of our grammatical custom that adds a doer to every deed.’ And as also pointed out in Lichtenberg’s (1959) objection to Descartes, it is not clear how one infers from mental states the idea of something that has those mental states. Consider again the case of a roll of thunder. When there is a thunder, there are various things happening that jointly give rise to the thunder. But one does not thereby conclude that there is something thundering. My purpose is to see if it is possible to adopt a similar way of thinking about mental states: one that will not force us into supposing that there is something, for example, that fears when a number of things happening jointly gives rise to fear.

Such a subject-free way of thinking about mental states is suggested by Parfit (1984) in his reductionist account of personal identity where he says ‘because we are not separately existing entities . . . we could fully describe our experiences, and the connections between them, without claiming that they are had by a subject of experiences’ (p.225). Parfit (1998) suggests that there could be imaginary intelligent beings who have no concept of themselves as the subjects of their experiences. Parfit (1998, p.221) claims that these beings are no worse than us, humans, both scientifically and metaphysically, that they think like us, have experiences like us, only without a conception of themselves as the subjects of their experiences. According to Parfit (1998), these beings have the concepts of persisting objects, of a sequence of thoughts, a sequence of experiences and acts, and they are capable of thinking that a particular sequence of thoughts occurs in a persisting body (p.228). Instead of thinking about what is involved in ‘seeing something’ or ‘feeling something’, they think of what is involved in ‘something being seen’, or a ‘seeing of something’, and instead of giving names to people, they give names to particular sequences (p.228-29). Hence where we say ‘Tenzing climbed Everest’, they say ‘in Tenzing there was a climbing of Everest’ (p.229).

These beings are also capable of making decisions, and are aware of their decisions, but they don’t think of themselves as making those decisions. Parfit
Thinking about Mental States

(1998) says that they conceive events like decision-making processes and the resulting acts as 'another kind of happening, distinctive only in the way in which these events are the product of practical reasoning, or, in simpler cases, of beliefs and desires' (p.229). As a result, a mountaineer would have thoughts like; ‘Should this include a crossing of that ridge of ice? Yes it should. And, unless that crossing starts now, it will be too late. So let the ascent begin!’ (Parfit, 1998, p.230)

The token condition would be undermined by Parfit’s suggestion of the possibility of such beings since these beings would have experiences without a conception of themselves as the subject of those experiences. Kriegel (2004) is willing to accept that it may be possible to think about a particular mental state without thinking about the subject whose mental state it is, but he contends that it would be abnormal for humans to think in this way. Obviously humans do have a conception of themselves as the subjects of their conscious mental states, but it may be more accurate to say that they have formed a conception of themselves as the subjects of their conscious mental states. Damasio (1999) has a detailed account of how the concept of oneself comes after consciousness during the course of the development of the mind. So, in thinking about one’s mental states, perhaps one naturally comes to think about oneself. But the token condition says something much stronger than this, viz. that it is impossible for a person to think about her particular mental state without thinking about herself. I cannot conceive any reason other than an extreme loyalty to our linguistic habits and a conceptual scheme arising from those habits why humans can’t think in this subject-free way or why Parfit’s beings would lead psychologically impoverished lives.

Perhaps it is impossible to think about a particular mental state without thinking about something else, but neither linguistic habits nor conceptual schemes and relations restricted in scope by linguistic habits should be determining what that something else is, especially if there is a theory of consciousness that eliminates the need for a subject to take the place of that something else. In what follows, I suggest that there is.

Even though Rosenthal endorses the token condition, I contend that his higher-order thought – henceforth I refer to this as HOT after Rosenthal (1997, p.741) – theory of consciousness or rather a slightly different version of it suggests an understanding of conscious mental states according to which it is possible to think about one’s mental states without thinking about oneself.

According to HOT theory, a mental state $x$ is conscious when there is another mental state viz. a thought about it. Mental state $x$ is typically described as the target state and the thought about it as the higher-order state. Rosenthal describes the thought as a ‘thought to the effect that one is oneself in that target state.’ (Rosenthal, 2005, p.129) Thus, in having a thought about my mental state, I also think about myself. However, as mentioned before, I don’t need to think about myself in a reflective, attentive manner. Although the description of the thought seems to be implied by the token condition, I have two remarks to make that might undermine the implication.
Firstly, although the content of the thought involves oneself, the higher-order state does not need to be conscious for the target state to be conscious. Hence, even if I am thinking about myself while thinking about a particular mental state that I am in, my thought about myself does not have to be conscious. However this does not immediately undermine the token condition. In favor of the token condition, one might argue that when the target state is conscious, the higher-order state nevertheless includes a thought about myself albeit an unconscious one, waiting as it were, to be phenomenologically activated when the higher-order state itself becomes conscious. But it should be noted that having an unconscious thought about oneself is phenomenologically no different from there being no thought about oneself. And as mentioned in the historical background, the idea behind the token condition typically feeds on this phenomenological difference i.e. on ‘… experiencing my experiences as mine.’ (Flanagan, 1992, p.194)

Still, it is worth considering whether the content of the higher-order thought can be described in another way. For instance, there is no reason to think that the higher-order state might not be an indexical thought with the content that state rather than I am in that state. As such it would be possible to re-describe the higher-order state in a more Humean or Parfit’s reductionist way. The purpose of the indexical type of higher-order state may be to assign the target state to a group of other psychologically continuous states – or if one favors bodily continuity theories of identity, to a group of bodily states that fall on the same spatio-temporal path – including the higher-order state itself. Hence, self-friendly content of the higher-order state which is I’m in so and so state may be replaced by the following self-free content of the higher-order state: this target state goes with this body and this series of psychophysical states. And when the higher-order state is conscious, one might then inferentially form the conception of oneself as the subject of a group of psychologically continuous states.

Nevertheless, there may be other reasons for the token condition. The first reason for the token condition, viz., the what-it-is-likeness is an essential part of experiences. Without the subject’s viewpoint as a part of the concept of an experience, thermometers can be said to experience heat and become hot or trees can be said to experience the presence of water or lack thereof and become thirsty. But we typically do not take trees to be thirsty and the thermometer to be hot. There is no what-it-is-likeness for the tree or for the thermometer.

The concept of an experience essentially includes what-it-is-likeness. And what-it-is-likeness involves a subject for which it is going to be like whatever it is like. So to undermine the token condition, one must give an account of the what-it-is-likeness without referring to the subject. I believe that one does not have to look far to see how this might be possible because it is something that is already suggested – but most of the time overlooked – by our very own concept of experiences.
Thinking about Mental States

Typically we attribute experiences to beings with a sufficient level of biological complexity, e.g. to those that have well-developed nervous systems (Damasio 1999). That is, we don’t think machines have experiences, but persons do. We don’t think unicellular organisms have experiences, but dogs and cats might. We don’t think flowers and trees have experiences but dolphins and chimpanzees probably do. This suggests that there might be a way to understand the what-it-is-likeness in virtue of an increasing level of biological complexity; i.e. in terms of interactions between simpler mental states, which give rise to experiences, and then in virtue of experiences themselves. If there is such a possibility, maybe we can understand the what-it-is-likeness of experiences without the subject that it-is-like-for. In other words, although the unique phenomenology of an experience is typically anchored in the subject’s viewpoint, there might be another way to make sense of the unique phenomenology of an experience that does not necessarily call for the subject.

Consider a belief that it will rain at time $t_n$. Consider further that this belief arises as a result of an observation of the sky at time $t_{n-1}$. A belief that it will rain at $t_n$ is associated with the perception of the sky at time $t_{n-1}$ along with many other psychophysical states. Among these are past sky-observations, some of which were followed by confirmations, i.e., observations that it did rain, and by disconfirmations, observations that it did not rain which eventually lead to the belief that it will rain at $t_n$ only when the sky-observation is in a particular way at $t_{n-1}$ based on previous experiences.

Likewise consider a present concern about a dentist visit that will take place in the future. It is a consequence of prior dentist visits, stories about dentist visits, a particular tolerance of pain, which has the threshold it has as a consequence of prior pains, knowledge of the particular procedure that is to be applied during the dentist visit or lack thereof that there is a concern in the first place and has the particular phenomenal character that it has.

This way of thinking about the unique phenomenology of experiences requires denial of the Humean type of atomism about experiences according to which ‘distinct perceptions are distinct existences’ and they are ‘separable from each other, and may be separately consider’d, and may exist separately, and have no need of any thing to support their existence’ (Hume, 1739-40/1978, p. 252). I assume that experiences presuppose relations to other experiences. What-it-is-likeness can then be said to consist not in the subject’s viewpoint but in the fact that each experience is associated with a unique relational matrix of other experiences. In other words, an experience is the experience it is and has the unique phenomenology it has as a consequence of the particular relational matrix of experiences it is a part of.

It would be interesting to study what types of interactions are at work in the relational matrix. However, I do not attempt to articulate the specifics of the relational matrix here. Perhaps the interactions are of a causal nature or maybe they are deterministic or maybe probabilistic. It may be suggested that the uniqueness consists in the particular spatio-temporal path of a person’s life. These are interesting questions but impossible to undertake within the scope of this paper.
Given the discussion above I suggest that it is impossible to think about a particular mental state token in isolation because I deny an atomistic view of experiences and not because I cannot think about a mental state without a subject. As such, subjects may be dispensable with regard to explaining the what-it-is-likeness feature of experiences.

Asymmetry in the Token Condition

Still, some may not be convinced by the reasons given above for the dispensability of subjects and endorse the token condition. However, even if the token condition is granted, and one agrees that one cannot think about a mental state token in abstraction from its subject, it is possible that this restriction applies only to the way we think about others’ and not about our own mental states. Consequently I suggest that there is an asymmetry between the way I think about my mental states and the way I think about another person’s mental states. I contend that given this asymmetry, the token condition does not apply to all mental state tokens.

As mentioned before, thinking about a particular mental state token implies thinking about what it would be like for someone to have it. In order to think about a particular mental state token of say the olfactory sensation of recently mowed grass, one needs to think of the olfactory sensation as somebody’s sensation. But when I think about a particular mental state of mine, its unique phenomenology is immediately given to me. I don’t need to further think about the subject whose mental state it is in order to understand that it has a unique phenomenology. So perhaps the special access a subject has to her mental states consists precisely in the fact that she does not need to further think of them as belonging to her.

Given the asymmetry described above, I contend that the way we think about a particular mental state token does not necessarily provide the right insight into the way we think about our mental state tokens. This is most likely a consequence of the gap between the immediate and non-inferential knowledge one has concerning her own (conscious) mental states as opposed to the inferential knowledge another has concerning those conscious mental states.

One might be curious about the nature of the gap involved. For my friend to know about the olfactory sensation that I’m having, she either needs to be told by me or she needs to observe me take in deep breaths with a pleasant look on my face and also realize that the grass has been recently mowed etc. I, on the other hand know it immediately. As mentioned by Kriegel (2004, p.198) too, there are no intermediate steps between my having the olfactory sensation and my knowing that I’m having such a sensation. This constitutes essentially an epistemological gap between my knowledge of my conscious mental states and someone else’s knowledge of them. This gap in turn gives rise to an asymmetry within the token condition: Perhaps it is impossible to think about
Thinking about Mental States

someone else’s mental state tokens without thinking about her but there is no good reason to think that the same restriction applies to my thoughts about my particular mental states.

Conclusion

My purpose in this paper was to see if there is a way to undermine the token condition, viz. the thesis that it is not possible to think about a particular mental state token without thinking about its subject. I suggested three reasons to reject the token condition; the first consisted in the weakening of the conceptual relation between experiences and their subjects; the second consisted in the possibility for a slightly different version of the higher-order thought theory of consciousness and the third consisted in the possibility of giving an account for the what-it-is-likeness of experiences without relying on the subject’s point of view. In the end I argued that even if the token condition is endorsed, the asymmetry between the way I think about my mental states as opposed to others’ mental states, suggests that the token condition does not apply to all mental state tokens.

References


C H A P T E R  T W E L V E

Naive Realism and the Explanatory Gap

Takuya Niikawa, Hokkaido University, Japan

To block the argument from hallucination, some naïve realists accommodate the eliminativist claim that hallucinations do not have visual phenomenology. It is, however, little known that the eliminativist claim also confers an ontological advantage on naïve realism. Taking into account the ontological advantage, I will argue, naïve realism can best address the explanatory gap problem.

Keywords: Naïve Realism, Explanatory Gap, Phenomenal Character,

Introduction: Eliminativistic Naïve Realism

Naive realism is a philosophical position about the phenomenology of veridical visual experience. It can be characterized as the conjunction of two claims, one explanatory and one metaphysical. The explanatory claim is that a veridical visual experience has phenomenology in virtue of the subject’s perceiving environmental objects, rather than his visual system representing such objects or his sensing private mental entities. An experience has phenomenology if and only if there is something it is like to have the experience. This explanatory claim does not say anything about the metaphysics of visual phenomenology. Given this, naïve realists are led to add the metaphysical claim that the phenomenology of veridical visual experience is wholly constituted by perceived objects and their properties.

There is a famous counterargument for naïve realism: the argument from hallucination. This argument shows that if we regard phenomenal aspects of veridical perception and corresponding total hallucination as metaphysically the same (common phenomenal factor principle), then naïve realism is untenable. Since the phenomenal aspect of hallucination cannot be constituted by environmental objects, the naïve realist metaphysical claim is incompatible with the common phenomenal factor principle. Given this, naïve realists must hold that the phenomenal aspects of veridical perception and hallucination are explained in different manners (phenomenal disjunctivism). What account should naïve realists provide of the visual phenomenology of hallucination? As is well-known, naïve realism has difficulty explaining the phenomenology of hallucination, due to the “screening off problem” illuminated by Martin (2004, p. 46). The problem is as follows: if the visual phenomenology of a

---

1For the argument from hallucination, see Fish (2009, pp. 29-33).
hallucinatory experience is sufficiently explained by a property P which a veridical experience also has, then it seems inevitable that the visual phenomenology of the veridical experience will be likewise explained by P. But, in this case, the perception relation, which naïve realists regard as explaining the visual phenomenology of veridical experience, seems to be explanatorily redundant or screened off, for all explanatory work seems to be performed by the common property P. If the perception relation is explanatorily useless, then naïve realism collapses. Hence, naïve realists need to devise an account of the visual phenomenology of hallucination that avoids this problem.

One bold strategy to block this counterargument is to accommodate the eliminativist claim: hallucinations do not have visual phenomenology. I call the conjunction of naïve realism and this eliminativist claim “eliminativistic naïve realism (ENR)”. If hallucinations do not have visual phenomenology, then naïve realists do not have to provide an account of the visual phenomenology of hallucinatory experience. Hence, there is nothing to screen off the explanatory power of the perception relation. Since ENR denies the need to explain the phenomenology of hallucination, it is more radical than phenomenal disjunctivism.

However, the eliminativist claim seems to be counterintuitive, implausible and even obviously wrong. It may be claimed that it is a pre-theoretical fact that hallucinations have visual phenomenology, or that the concept of hallucination necessarily involves the possession of visual phenomenology. In response, Fish (2009, section 5) and Logue (2012) argue that ENR is, at the very least, coherent. Although it is controversial whether their arguments have succeeded, I assume the coherence of ENR in this paper.

In addition to blocking the argument from hallucination, the eliminativist claim confers an ontological advantage on naïve realism. Little attention has been given to this positive aspect of the eliminativist claim. The aim of this paper is to argue that ENR is the best theory to deal with the explanatory gap problem. Roughly speaking, my argument is as follows. William Fish (2008; 2009; 2013) argues that naïve realism alone can dissolve the explanatory gap problem, which has been illuminated by Joseph Levine (1983). However, Adam Pautz (2013) makes two objections to Fish’s argument. The first is that naïve realism can at best transform the explanatory gap problem, rather than dissolving it. The second is that not only naïve realism but also representationalism can transform the explanatory gap problem in the same manner. In response to the first objection, I will argue that such a transformation is epistemologically desirable. For the second objection, I will argue that naïve realism acquire an ontological advantage over representationalism by taking in the eliminativist claim.

Two Explanatory Gaps

The explanatory gap problem, to which Fish (2008; 2009) claims that
Naive Realism and the Explanatory Gap

Naïve realism can provide a solution, is as follows. On a standard qualia theorist view, the phenomenology of a visual experience is an intrinsic, non-representational property of the subject, which nomologically supervenes on certain neural processes in the subject’s brain. Suppose that you are seeing a red apple. In this case, you are likely to have a visual experience with reddish-apple-phenomenology. The qualia theorist maintains that the reddish-apple-phenomenology nomologically supervenes on certain neural processes in your brain. The question to be asked here is, why is the neural processes connected with the reddish-apple-phenomenology, rather than bluish-apple-phenomenology or reddish-banana-phenomenology? We seemingly cannot fully explain, by appealing to any functional or physical properties of the neural processes, why you are undergoing the reddish-apple-phenomenology, rather than some other type of phenomenology. This indicates that there is an explanatory gap between the phenomenology of an experience and the neural processes responsible for the experience.¹

Fish (2008; 2009; 2013) argues that naïve realism can close the explanatory gap. According to Fish, “it is the nature of this environment [the subject is acquainted with]–not the nature of the [brain] processing–that accounts for what it is like to [have a visual experience]” (2013, p.58). That is to say, it is not neural processes but environmental objects and their perceptible properties that we should appeal to in order to answer the question of why you are undergoing the reddish-apple-phenomenology rather than some other type of phenomenology. Since naïve realism holds that the reddish-apple-phenomenology is constituted by the perceived apple and its redness themselves, it does not make sense to further ask why and how the perceived apple and its redness constitutes the reddish-apple-phenomenology, rather than some other type of phenomenology. Such a question is similar to the following nonsense question: why and how do these iron parts constitutes the iron statue, rather than some other type of statue, such as a wood statue? A person who asks such a question is merely misunderstanding the concept of constitution. On the naïve realist view, therefore, there is no explanatory gap between the phenomenology of veridical experience and the perceived environmental objects. If it is inappropriate to explain in terms of certain neural processes why you are undergoing the reddish-apple-phenomenology rather than some other type of phenomenology, then there is no gap between the reddish-apple-phenomenology and the neural processes that we should close by adding some explanatory story. On Fish’s proposal, therefore, the explanatory gap problem is not solved but is dissolved. Such a problem does not exist from the beginning.

Pautz (2013) plausibly claims that what naïve realism can indeed do is not to close the explanatory gap but simply to replace it to a different position,

¹This explanatory gap should be distinguished from a different explanatory gap, which is characterized in terms of the question of why there is something it is like for you to see a red apple. This question is distinct from the question at issue: why is the visual experience of a red apple connected with reddish-apple-phenomenology rather than some other type of phenomenology? What naïve realism can answer is not the former question but the latter one.
which is between microphysical properties (e.g. particular reflectance and arrays of particles) and perceptible properties (e.g. redness and apple-shape). It is undeniable that environmental objects have not only perceptible properties but also microphysical properties. Moreover, it seems plausible that an object has a perceptible property in virtue of having certain microphysical properties. Given these, suppose that your experience has reddish-apple-phenomenology. It is not explanatorily enough to say that this is because you are seeing a concrete physical object, and its redness and its apple-shape constitute the phenomenology. We should further explain why and how the particular perceptible properties (redness and apple-shape), rather than other perceptible properties (say, greenness and banana-shape) are connected with the microphysical properties of the object. This explanatory task seems no less difficult than the original one.

In response, Fish (2013) claims the gap should be located between the microphysical properties and perceptible properties (I call this the “objective gap”), rather than between neural processes and intrinsic, non-representational properties (I call this the “subjective gap”). But in what sense is the objective gap better than the subjective one?

One possible way to answer is to emphasize that the objective gap is indeed “original”. As Shoemaker (2003) has explicated, the explanatory gap was located at the objective position before dualism of mind and matter flourishing, but it has been moved to the subjective position by dualists. In this light, it may be claimed that the objective position is the good-old one and that the gap should be moved again by non-dualists to the original position. It is likely that Fish follows this line, because he writes:

The reason we find a subjective explanatory gap in the first place is because it was hoped that the objective explanatory gap could be dealt with by kicking it upstairs into the mind. That might work for a dualist, but for a materialist the problem simply resurfaces in a new, subjective, form. So in relocating it to the world, I am simply putting it back in its original place. (2013, p. 58)

However, it is unclear why the gap should be located at the original position for non-dualists. The fact that an item was originally located at a position does not by itself mean that the item should be there. Fish adds this claim:

if we could make it plausible that […] the character of a particular episode of acquaintance arises from the environment the creature is acquainted with, then we would have made great strides in understanding consciousness, even if there did remain an unsolved problem in metaphysics. (2013, p.58)

His point is that if the gap is moved again from the subjective position to
the (original) objective position in the naïve realist manner, we can obtain a better understanding of perceptual consciousness. However, it is hard to see why that is so. The subjective gap problem consists in the fact that it is unclear why a particular intrinsic property of a subject, rather than some other intrinsic property, is instantiated by certain neural processes. This problem is related to perceptual consciousness, precisely because the intrinsic property is regarded as identical to the phenomenology of perceptual experience. On the other hand, the objective gap problem consists in the fact that it is unclear why a particular perceptible property of an object, rather than some other perceptible property, is realized by certain microphysical properties. Note that naïve realists accept the phenomenal externalist idea that (veridical) perceptual consciousness essentially involves environmental objects and their perceptible properties. Given this, the objective gap problem also seems to be related to perceptual consciousness, because perceptible properties are not ontologically distinct from the phenomenology of veridical experience in the naïve realist framework. As long as we hold on to the phenomenal externalist idea, the question of why a particular perceptible property is realized by certain microphysical properties—“an unsolved problem in metaphysics”—is regarding perceptual consciousness. As we have seen, the objective gap problem is no less difficult than the subjective one. Qualia theorists arguably contend that phenomenal consciousness is located in the brain; naïve realists insist that phenomenal consciousness spread beyond the brain to the environmental world. The fact that these conceptions lead to different explanatory gaps, by itself, does not make us favor one over the other.

Nevertheless, there is at least one argument to show that the subjective gap is more harmful than the objective one from an epistemological perspective. On the qualia theorist view, the phenomenology of a visual experience is private in the sense that a particular token of an intrinsic property is manifested only to the experiencing subject. In this case, we cannot know, via the direct observation of the token, what phenomenology others are undergoing. Here, an epistemological problem arises: it is unclear how we can know what phenomenology others are undergoing. The problem partially stems from the subjective gap. The gap can be plausibly extended from the explanatory level to the metaphysical level. This opens up the metaphysical possibility that the same functional or physical type of brain activity realizes radically different intrinsic properties. Given this possibility, the fact that our brain activities are significantly similar in functional or physical properties does not entail that we have similar phenomenology. Hence, even if you and I see significantly similar scenes and we are almost the same in neural processes, I might undergo radically different phenomenology from yours. This seems to destruct the commonsense idea that the phenomenology which I am undergoing in a perceptual situation is reliable evidence regarding what phenomenology others undergo in a similar situation. Arguably, this leads to the skepticism that we cannot know what phenomenology others undergo.

In response, it may be claimed that we do not need to rule out the above metaphysical possibility in order to know what phenomenology others
undergo. Suppose that there is a *nomological* correlation between the phenomenology of a type of veridical experience and certain neural processes in this actual world. Given that neural processes of those who are in a similar perceptual situation are similar to each other, we would inferentially get to know what type of phenomenology others are undergoing. However, it is unclear how to justify the presence of such a nomological correlation. Perhaps, I can verify the presence of such a correlation *in my case*. When I have a visual experience, I might be able to somehow observe the neural processes of my brain and then contrast them with the phenomenology of the experience. Based on the continuous observation, I might find out a nomological correlation between them. However, it is unjustifiable to generalize the correlation, unless we find out some mechanism to connect the phenomenology of visual experience with certain neural processes. Indeed, the subjective explanatory gap stems from the lack of such a mechanism in the first place.

In contrast, on the naïve realist view, the phenomenology of a veridical visual experience is objectively observable in the sense that we can directly observe a particular token of a perceptible property which constitutes the phenomenology of the veridical experience of others. Suppose that you and I are significantly similar in functional properties relevant to perception. On this supposition, when you and I are seeing the same red apple, we are undergoing the same type of phenomenology that is constituted by the apple and its redness. If I know that you are seeing the same apple as I am, and that the phenomenology of my visual experience is constituted by the apple, then it seems that I am able to inferentially know what phenomenology you are undergoing. Even if the objective gap can be extended to the metaphysical level, any epistemological problem does not arise. Since perceptible properties themselves are objectively observable, the fact that others are seeing a particular instance of a perceptible property is visually knowable. Given this, the metaphysical possibility that the same microphysical properties realize radically different perceptible properties does not lead to the skepticism that we cannot know, via direct observation, what perceptible properties others are seeing. The metaphysical possibility can cause a relevant epistemological problem only when all we can directly observe are microphysical properties. This is obviously not the case. Therefore, even if we take into account the metaphysical possibility, there is no reason to deny that we are able to know what phenomenology others are undergoing. This is the epistemological reason for favoring the objective gap.

Note that this is not to say that we can *perfectly* know what phenomenology others are undergoing. Taking into account our individual differences, this is very implausible. My point is rather that it is commonsense that we can know what phenomenology others are undergoing *to some extent*, and that the subjective gap is destructive to this commonsense idea. Suppose that I am short-sighted and you are clear-sighted. Even if we are seeing the same red apple, our visual experiences are different in phenomenology. The phenomenology of my experience is not constituted by the determinate shade
and shape of the apple, because I do not have a perceptual capacity enough to recognize such shade and shape. Nevertheless, since I can see the determinable color and shape of the apple, the phenomenology is constituted by the apple and its determinable color and shape. In contrast, the phenomenology of your experience is constituted by the apple and its determinate shade and shape. This entails that it is also constituted by the determinable color and shape, which constitute the phenomenology of my experience. In this case, I cannot perfectly know what phenomenology you are undergoing, simply because the constituents of my experience are not entirely the same as yours. Nevertheless, I can know what phenomenology you are undergoing to some extent, since our experiences have the common constituents, such as the determinable color and shape.¹ It is intuitively plausible that in everyday situations, we can know what phenomenology others are undergoing to such extent. As we have seen, the qualia theorist view has to let go of this commonsense idea. Because of this epistemological advantage, the explanatory gap should be located at the objective position rather than the subjective position.

**ENR or Representationalism**

Can naïve realism alone locate the explanatory gap at the objective position? The main rival of naïve realism is representationalism, which is characterized as a conjunction of the following explanatory and metaphysical claims: (1) a visual experience has phenomenology in virtue of the subject’s visual system representing environmental objects and their properties; (2) the phenomenology of a visual experience is identical to a certain representational content of the experience (I call the content RCE).² As Pautz (2013) has pointed out, some kinds of representationalism accept that “the sensible qualities are in the mind-independent world” (p.29). That is, representationalists can hold that RCE contains perceptible properties as its components. Given this, such representationalists can identify the phenomenology of veridical visual experience with perceptible properties which can be instantiated by environmental objects. This suggests that naïve realism and representationalism do not differ in where to locate the explanatory gap. However, this does not mean they are entirely on a par. In this section, I will argue that ENR is ontologically better than representationalism.

First, it should be noted that RCE is an abstract entity, which cannot be identical to a concrete spatiotemporal component of the environmental world. The question to be asked here is, what kind of abstract entity is RCE? Note that the phenomenology of veridical visual experience is presentational in nature in

---

¹For the naïve realist account as to how the environmental world feature in our phenomenal experience, see Fish (2009, section 3).

²There is also a weaker metaphysical characterization: the phenomenology of a visual experience supervenes on a certain representational content of the experience. This characterization is insufficient in that it does not indicate anything about the ontological status of visual phenomenology (Tye 2009). Since this paper focus on such ontological status, I adopt the stronger characterization.
the sense that a veridical visual experience phenomenally presents the subject with something that seems to be environmental objects. RCE must be explanatory of this presentational nature of veridical visual experience. I call this explanatory requirement the “presentational requirement”. Now, representationalists have two options. The first is to characterize RCE as a distinctive type of content, which is essentially different from the type of content attributed to cognitive activities/states, such as thinking, judging or beliefs (I call such content “cognitive content”). The second is to think that RCE is not essentially different from cognitive content. A type of content is essentially different from another type of content if and only if each type of contents is metaphysically analyzed in terms of a different kind of abstract entity. I will discuss this point in detail later. My argument is that either option leads to, in different manners, the same conclusion that naïve realism has an ontological advantage over representationalism.

According to the first option, RCE is essentially different from cognitive content. On this option, it seems that the presentational requirement can be satisfied, because representationalists can freely introduce a special type of content that is defined as satisfying it. It is not incoherent to think that RCE is primitively explanatory of the presentational nature of veridical visual experience. However, this option imposes an ontological burden on representationalists. In order to metaphysically analyze such special type of content, they must accommodate a distinct abstract entity in their ontology, in addition to the abstract entities to analyze cognitive content. It is of course not reprehensible to introduce an abstract entity in order to explain a phenomenon. My point is rather that we should be economical with respect to ontological commitment. That is to say, we should not introduce an abstract entity if an explanatory task at hand is achieved by reference to entities in hand. Naïve realists do not require any abstract entity for characterizing the phenomenology of veridical experiences. All they need to invoke are concrete physical objects instantiating perceptible properties, which representationalists must also accommodate in their ontology.

Perhaps, representationalists can establish a reductive theory of RCE in which RCE is fully analyzed in terms of scientifically admissible entities.¹ In this case, representationalism does not involve any additional abstract entities. However, there is a fairly general agreement that no satisfactory reductive theory has been established.² Unless such a theory is actually constructed, we can permissibly exclude it from our consideration. Thus, if representationalists take the first option, it follows that naïve realism is more ontologically economical than representationalism.

According to the second option, RCE is not essentially different from cognitive content. Representationalists who takes this option may claim that naïve realism does not have any advantage in ontological economy, since naïve realists should also admit that cognitive activities/states have representational

¹Dretske (2003) and Tye (2009) are the representative advocates of such a reductive theory.
²For a comprehensive objection to such reductive theories, see Pautz (2010).
contents. In response, naïve realists have a suspicion that it is difficult to satisfy the presentational requirement by appealing to the type of content that is shared with cognitive activities/states, since cognitive activities/states are not presentational. This is certainly true, but representationalists do not have to appeal to the *same* type of content as that of cognitive content. Even if two types of contents are not essentially different, it does not entail these types are the same. That is to say, even if two types of contents are metaphysically analyzed in terms of one kind of abstract entity, these types of contents can be differently characterized. Suppose that a type of content is characterized in terms of singular propositions and another type of content is characterized in terms of general propositions. In such a case, this characterization commits us to only one abstract entity, proposition. Nevertheless, each type of content would have different explanatory power.

There is, however, a significant difference between naïve realism and representationalism as to what abstract entity can be used to metaphysically analyze cognitive content. Naïve realists do not have to characterize visual phenomenology in terms of representational content. Hence, the abstract entities that are employed to analyze cognitive content do not need to have the potential to satisfy the presentational requirement. In contrast, representationalists must use RCE to characterize visual phenomenology. Hence, the abstract entities that are employed to analyze cognitive content must have the potential.

This difference indicates that naïve realism is open to more various metaphysical analyses of cognitive content as compared to representationalism. Take the possible world analysis for example.¹ Suppose that one believes that there is a red apple in front of him/her. On this analysis, the content is analyzed as the set of possible worlds in which there is a red apple in front of him/her (or his/her duplicate). Let us plausibly assume that satisfaction conditions of cognitive activities/states are adequately captured by this possible world analysis. Under this assumption, this analysis is available to naïve realists. However, representationalists seemingly cannot adopt it. The possible world analysis has not been intended to capture the presentational nature of visual experience in the first place. It is indeed unclear how a relation to a set of possible worlds can explain the phenomenology of a veridical visual experience. It seems extremely implausible that when one has a visual experience with the RCE that there is an apple in front of him/her, all possible worlds in which there is an apple in front of him/her are phenomenally presented to him/her. From the phenomenological perspective, it seems that a world (whether this is the actual world or a non-actual world) alone can be presented to us at a time. If this consideration is correct, then the possible world analysis cannot be applied to RCE, which is supposed to be identical to the phenomenology of visual experiences. Therefore, while naïve realists can include the possible world analysis in the candidate list of metaphysical analyses of cognitive content, representationalists cannot do so.

Generally speaking, the concept of representational content has been

---

¹For the detail of this analysis, see Stalnaker (1984) and Lewis (1986, section 1.4)
mainly used to capture the satisfaction conditions of mental states. Given this, it is reasonable to think that some other existing metaphysical analyses of cognitive content do not meet the presentational requirement. Thus, the second option imposes much constraint on the metaphysical analysis of cognitive content. This is a severe disadvantage of representationalism.

In response, representationalists may claim that the presentational nature of veridical visual experience should be explained not by RCE itself but the special mode of representing. Perceptual experience represents a certain content in a presentational manner; cognitive activity/state represents a certain content in a non-presentational manner. On this strategy, RCE itself does not have to meet the presentational requirement. However, this idea leads to the same problem as the first option. Since the relation of representing in a presentational manner is not included in the naïve realist ontology, naïve realism is ontologically more economical than representationalism.

Representationalists might maintain that naïve realism must include the special acquaintance relation in its ontology in order to explain the phenomenology of veridical visual experience, instead of the special representation relation. If so, naïve realism and representationalism does not differ in ontological economy. However, this is not the case. The explanatory claim of naïve realism has been that a veridical visual experience has phenomenology in virtue of the subject’s perceiving environmental objects. The perception relation should be understood in an ordinary sense, rather than in a unique sense used by naïve realists alone.1 It is almost undeniable that we can perceive environmental objects in such an ordinary sense. Hence, representationalists must accommodate the perception relation into their ontology. Thus, if representationalism draws upon the special representation relation, it follows that naïve realism has an ontological advantage over representationalism. Consequently, whichever option representationalists take, naïve realism has an ontological advantage over representationalism.

Most importantly, the eliminativist claim is crucial for this argument. If naïve realists dismiss it, then they need to accommodate certain entities (say, qualia, sense-data or a certain type of representational content) in order to explain the phenomenology of hallucinations. The accommodation of such entities obviously deprives naïve realism of the ontological advantage mentioned above. ENR alone can appeal to the advantage.

**Conclusion**

From the above discussion, we can reasonably conclude that ENR can best deal with the explanatory gap problem. This would count as an important motivation for naïve realism.

---

1If you cast doubt on the existence of the ordinary sense of “perceiving”, naïve realists can exchange it with “seeing”. There is likely to be no one denying the existence of the ordinary sense of “seeing”.
Acknowledgments

Thanks to anonymous referee, and Ichiro Nishida, Joel Smith, Koichi Nakashima, Tomoyuki Yamada, Yu Yoshii. Special Thanks to Tohru Genka and Yasushi Ogusa for their detailed written comments. I could refine section 3 based on their comments. I gratefully acknowledge the support of JSPS.

References

Donald V. Poochigian, University of North Dakota, USA

David Bohm identifies a paradox within quantum theory. Assumed by quantum theory is an observer who is both a distinguishable constituent of an indistinguishable world, and an indistinguishable constituent of a distinguishable world. Without a distinguishable observer, a distinguishable quantum world is impossible. Within a distinguishable quantum world, a distinguishable observer is impossible. Thus, a quantum world appears self-contradictory.

Content in relation composes a quantum world. A distinguishable and indistinguishable quantum world having the same content and different relation, content is not paradoxical and relation is paradoxical. Non-quantum and quantum worlds being ontologically incompatible, relational identity is an inherent operational and functional constituent of the quantum world. Component of relation are the operators one and zero, and the functions injection, surjection, and bijection. Integrating the minds of distinct scientific observers is the relativistic bijective assumption of common experiential content and uncommon experiential form. Quantum operational functions smoothly transitioning quantum states from one into another, contradictory states occur consecutively, not concurrently. Doing so resolves Bohm’s paradox.

Bohm’s Paradox

At the foundation of quantum mechanics is a paradox. While widely acknowledged among quantum theorists, David Bohm explicitly formulates it noting,

at the quantum level of accuracy the entire universe (including, of course, all observers of it), must be regarded as forming a single indivisible unit with every object linked to its surroundings by indivisible and incompletely controllable quanta. If it were necessary to give all parts of the world a completely quantum-mechanical description, a person trying to apply quantum theory to the process of observation would be faced with an insoluble paradox. This would be so because he would then have to regard himself as something connected inseparably with the rest of the world. On the other hand,
the very idea of making an observation implies that what is observed is totally distinct from the person observing it (Bohm 584-585).

Constitutive of this enigma is assuming investigation of the natural world constitutes statistical correlation of observational experiences. Accepting this, then to what investigation of the natural world comes is quantum theory. Statistics is ontologically presented as a quantum entanglement linking empirical observations. Composed is an unbroken continuity of combinations of initiatory and concluding observations. Concluding observation is a ‘collapse’ of the quantum entanglement in this otherwise unbroken continuity.

Presuming attainable is, ‘As long as Natural Philosophy exists, its ultimate highest aim will always be the correlating of various physical observations into a unified system, and, where possible, into a single formula,’ then everything in the universe is related (Planck 1). Accepting a relation is a quantum entanglement of quantum collapses, since a quantum entanglement of any two quantum collapses, $a$ and $b$, is imaginable, then everything in the universe is related. A quantum entanglement constituting every state of being intervening $a$ and $b$, it is an unbroken continuum from $a$ to $b$. Now $a$ and $b$ are indistinguishable as autonomous objects. They are distinguishable only as states of being of a common object, $ab$. Metaphorically constituted is,

A rhizome . . . made of plateaus. . . . We call a “plateau” any multiplicity connected to other multiplicities by superficial underground stems in such a way as to form or extend a rhizome. . . . Each plateau can be read starting anywhere and can be related to any other plateau. To attain the multiple, one must have a method that effectively constructs it (Deleuze&Guattari 21-22).

Even this ‘Babylonian attitude [where] you know all of the various theorems and many of the connections in between,’ is incomplete, considering every quantum entanglement is composed of an infinity of points (Feynman Chapter 2, “The Relation of Mathematics to Physics”). Each of these points is linked to every other point in a quantum entanglement, until generating an undifferentiated line. In turn, each quantum entanglement is linked to every other quantum entanglement in a quantum entanglement, until generating an undifferentiated surface. Each undifferentiated surface is linked to every other undifferentiated surface in a quantum entanglement, until generating an undifferentiated universe.

Content of the universe being undifferentiated, neither points nor quantum entanglements are discernable within it. Neither points nor quantum entanglements being discernable within the universe, then paradoxical is quantum generation of the universe. Simply enough, there are no elements by which to generate the universe. Thus, on its own assumptions, quantum theory is self-contradictory.

Everything within the universe being related, nothing within the universe is distinguishable except as a state of being of the universe at quantum
collapse. Nothing being distinguishable as independent of the whole, neither observer nor observed is distinguishable. Observer and observed defining observation, when observation defines quantum collapse, then quantum collapse is impossible. Science being observation at quantum collapse, then science is impossible. Now as ‘correlating of various physical observations into a unified system,’ science is possible when it is impossible. This is a paradox.

**Classical Solution**

Bohm seeks to resolve the conundrum arguing, ‘The paradox is avoided by taking note of the fact that all real observations are, in their last stages, classically describable’ (Bohm 585). Doing so confronts Sir Isaac Newton’s,

Law I.

*Every body perseveres in its state of rest, or of uniform motion in a right line, unless it is compelled to change that state by forces impressed thereon.* . . .

Law II

*The alteration of motion is ever proportional to the motive force impressed; and is made in the direction of the straight line in which that force is impressed* (Newton 19, 20).

Concerning ‘a right [straight] line,’ at the limit of a one-dimensional extension is an infinity of immediately contiguous points. Extension to any one of these points is no more necessary than extension to any other. Excluding all points but what Newton might identify as ‘uniform motion in a right line,’ this could be at a juncture of two equally offset points. Now progress ‘in a right line’ is ambiguous, constituting a diagonal movement onto one or the other offset points. Additionally, being arbitrary, neither movement is necessary. Progressing to one or another offset point, ambiguity recurs in a new form. Is ‘uniform motion in a right line,’ continuation ‘in a right line’ of the diagonal away from what initially would have been ‘uniform motion in a right line?’ Or is it away from continuation ‘in a right line’ of the diagonal, in a diagonal returning towards what would initially be continuation ‘in a right line?’

Again, what constitutes ‘in a right line’ is ambiguous. As ambiguous, extension to either point is no more necessary than extension to the other. Thus, Newton is mistaken in thinking motion in the physical world is necessary. Rather it is unnecessary, and being unnecessary, what constitutes motion in the physical world is probable. Being so, quantum theory is affirmed. Now God does play dice with the universe.

Probability being intrinsic to the geometry of the physical universe, ontologically suspect ‘forces impressed upon [Every body]’ are redundant. However ambiguity at the limit of a one-dimensional extension of an object is
resolved, ambiguity can recur at every limit of a one-dimensional extension. Being so, the path of any physical object is rendered uncertain.

**Quantum Solution**

Quantum theory being reaffirmed, Bohm’s paradox still is not avoided. For not only is the path of any physical object uncertain, constitution of the path of any physical object is uncertain. Path of a physical object is a continuum from initial object \( a \) to subsequent object \( b \). How is it known, though, initial object \( a \) is subsequent object \( b \)? Ernst Zermelo proposes a solution in the axiom of choice.

Among the representations of the axiom of choice is,

**AXIOM VI. Axiom of choice (Axiom der Auswahl):** “If \( T \) is a set whose elements all are sets that are different from [the null set] and mutually disjoint, its union \( UT \) includes at least one subset \( S_1 \) having one and only one element in common with each element of \( T \)” (Wikipedia ‘Zermelo set theory.’ Also see Barker 77; Bernays 133; Fränkel 16).

Not indicated is how it is known \( S_1 \) is the same ‘element’ in each of the subsets of \( T \). If \( T \) is identified by its relations, then \( S_1 \) is not the same element in each of the subsets of \( T \) because it has different relations in different subsets.

Additionally requiring ‘at least one subset \( S_1 \) having one and only one element in common with each element of \( T \)” is problematic because being ‘at least’ so. Since there can be more than one such subset: \( S_1, S_2, \ldots S_n \), how is \( S_1 \) distinguishable from \( S_2 \ldots S_n \)? After all, each has the same properties as every other.

Difficulty occurs because the axiom of choice identifies \( S_1 \) by its relational properties. \( S_1 \) is distinguishable from \( S_2, \ldots S_n \), and \( S_2, \ldots S_n \) from each other, only by what uniquely identifies them. This is membership in a set of one, which constitutes haecceity, the property of being oneself and nothing else.

As such, though, neither \( S_1 \) nor \( S_2 \ldots S_n \) are constituent of \( T \). Now there is no necessity in identity of observed object \( S_1 \) and observed object \( S_2 \). Neither is there any necessity in every subsequent transitive identity \( S_3 \ldots S_n \) of \( S_1 \) and \( S_2 \).

It can be replied how element is related to element is unimportant, because in quantum theory everything is related in entanglement. All being entangled, however, how is anything distinguishable? This when science is the identity of relationship between distinguishable things.

Confusion arises because quantum theory is a contradictory axiom system composing many and one, without an axiom of resolution. Introducing an axiom of resolution by ‘Explicit enumeration’ provides consistency (Blanche 22). Occurring by ‘Explicit enumeration,’ an axiom of resolution is injected from outside the quantum world.
Metaphysical Solution

Incompleteness of quantum theory in respect to an axiom of resolution is why, ‘No way is evident to apply the conventional formulation of quantum mechanics to a system that is not subject to external observation.’ (Everett 455). Concerning the character of this requisite ‘external observation,’ John von Neumann concludes,

it is inherently entirely correct that the measurement or the related process of the subjective perception is a new entity relative to the physical environment and is not reducible to the latter. Indeed, subjective perception leads us into the intellectual inner life of the individual, which is extra-observational by its very nature (since it must be taken for granted by any conceivable observation or experiment) (von Neumann 418).

Acknowledging the ‘intellectual inner life of the individual,’ Erwin Schrödinger observes, ‘The abrupt change by measurement . . . is precisely the point that demands the break with naive realism’ (Schrödinger “7. The Psi-function as Expectation-catalog”). So being, ‘Resolution of the “Entanglement” Result [is] Dependent on the Experimenter's Intention’ (Schrödinger “11. Resolution of the ‘Entanglement’ Result Dependent on the Experimenter's Intention”). Thus, ‘Which measurements on B and in what sequence they are undertaken, is left entirely to the arbitrary choice of the experimenter’ (Schrödinger “11. Resolution of the ‘Entanglement’ Result Dependent on the Experimenter's Intention”).

Summarizing the observer’s role, Hugh Everett concludes,

Throughout all of a sequence of observation processes there is only one physical system representing the observer, yet there is no single unique state of the observer . . . there is a representation in terms of a superposition, each element of which contains a definite observer state and a corresponding system state (Everett 459).

Here Everett is understandable as supposing a metaphysical conception of mind, rather than a physical conception of mind as brain, as in many other versions of the many minds variation of quantum theory.

Building on Everett’s supposition, David Albert and Barry Loewer operationalize Everett’s conception with a formal account of mind and brain (Albert&Loewer 195-213). Insofar as, ‘Minds do not obey the Schrödinger evolution (in particular the superposition principle), but evolve in time in a genuinely probabilistic fashion,’ minds are not axiomatic (Hemmo&Pitowsky 133-176). They are constituted of neither enumerated elements, nor enumerated sequences of elements. Insofar as minds ‘evolve in a genuinely probabilistic fashion,’ they evolve indeterminately.
Brains are axiomatic, containing enumerated elements and enumerated sequences of elements. Assuming a teleological axiom system, brain states do evolve in a probabilistic manner. Difference in genuine and axiomatic probability is genuine probability is unlimited, and axiomatic probability is limited. Thus, genuine probability includes all occurrent elements and sequences, while axiomatic probability excludes all unenumerated elements and sequences. Genuine and axiomatic probability are integrated by being definitional conceptions of the same elements, but are separated by being different definitional conceptions of the same elements.

Consciousness

Reducing mind to brain engenders its own paradoxes, psychological materialism suffering from unresolvable difficulties. Denying consciousness, psychological eliminativism is self-contradictory, claiming awareness of there being no awareness. Accepting consciousness, psychological reductivism and property dualism are circular, consciousness a Frankenstein monster creating the matter that is consciousness’ own creator.

John R. Searle (Searle 133-176) and David Chalmers (Chalmers 243-244, 255) seek to evade this circle by viewing one side of it from the other side. Searle criticizes Chalmers by explaining consciousness from the perspective of a constituent of consciousness, the propositional attitude of a real autonomous material world. Chalmers criticizes Searle by explaining a constituent of consciousness, the propositional attitude of a real autonomous material world, from the perspective of consciousness. Resolution is accepting the circularity. Each perspective incomplete because explained by the other, both perspectives are complete when understood as mutually explaining each other.

In dispute is the reality of relationship. Causality being a relationship, matter can cause mind only if relationship is real, existing independently of mind. Relationship being unreal, mind constitutes matter as mind’s own cause. Accepting the latter, reality is pan-psychic, relationship imposed by mind. Accepting the former, reality is not pan-psychic, relationship received by mind. Science seeks a correspondence between hypothetical appearance and actual reality.

Escape from this conclusion might be sought by arguing consciousness is hypothetical, a propositional attitude, not an observable entity, because all could be real as well as apparent. Consciousness as hypothetical is self-contradictory, though. Being analytic, consciousness as hypothetical presupposes consciousness as manifesting appearance.

Consciousness either is or is not, it is not hypothetical, not something which might or might not be. Because so, matter as the source of consciousness is circular. Since this sourcing is unobservable, it is hypothetical. Being hypothetical, its supposition is a creation of consciousness. Now consciousness as the created is the creator of its own creator.
Bohm's Paradox and the Conscious Observer

More generally, neither body nor mind being occurrent in experience, when same experience can be the product of body or mind, body and mind are ascribed abstractions. Experience is real when ascribed to body, and ideal when ascribed to mind. Ascription being constituent of consciousness, consciousness is actual, and body and mind are nominal.

**Psycho-Physical Parallelism**

Insofar as it, is a fundamental requirement of the scientific viewpoint – the so-called principle of the psycho-physical parallelism – that it must be possible so to describe the extra-physical process of the subjective perception as if it were in reality in the physical world, then mind as an autonomous axiom system must compose physical objects (von Neumann 419). So being, mind cannot avoid Bohm’s paradox. Insofar as, ‘no definition of what an “observer” is[, is] available, in terms of an atomic scale description, even in principle,’ a unified system of physical observations is unattainable (Price ‘Q31 What is the Copenhagen interpretation?’).

Only by introducing metaphysical mental operators, the ‘intellectual inner life of the individual,’ is Bohm’s paradox avoided. This is, ‘to [not] describe the extra-physical process of the subjective perception as if it were in reality in the physical world.’ Assuming, ‘the entire universe . . . must be regarded as forming a single indivisible unit with every object linked to its surroundings by indivisible and incompletely controllable quanta,’ then the axiom system depicting the observer must be ‘connected inseparably with’ the axiom system depicting the observed. As such, neither is distinguishable from the other, Albert and Loewer’s many minds variation of quantum theory failing.

Escape from this conundrum is possible only when embracing ‘the extra-physical process of the subjective perception,’ (von Neumann 419) and abandoning proceeding ‘as if it were in reality in the physical world.’ Certainly, doing so renders ‘Natural Philosophy’ incomplete, and thus incapable of achieving ‘a unified system.’ Not doing so, though, renders ‘Natural Philosophy’ self-contradictory.

Embracing ‘the extra-physical process of the subjective perception,’ however, is challengeable on the ground of incomparability. Insofar as consciousness is metaphysical, and matter is physical, they are ontologically incommensurable. This constitutes ‘the hard problem,’ which is the foundation of psycho-physical parallelism. Relation is possible, however, considering although ontologically incommensurable, consciousness and matter are epistemologically commensurable.

They are so insofar as consciousness is intensional. This is when constituent of consciousness is a sense of identity with an object. So occurring, mind as consciousness identifies with body when as constituent of mind as
consciousness is identity with body. Metaphorically, consciousness is like a transparency overlaying matter, and thus ordering matter, while not interacting with matter.

**Micro Functional Operators**

This ordering proceeds by distributing ambiguous members of a set. Such members being identifiable, together they constitute the subset of all mutually ambiguous constituents of the set. These elements being mutually identifiable as constituent of this subset, their membership in this subset is unambiguous. Limitless application of inclusive disjunction to the subset of all ambiguous members of two sets defines mathematical one (‘1’). Limitless application of exclusive disjunction to the set of all ambiguous members of two sets defines mathematical zero (‘0’).

Resolving disjunctions inclusively or exclusively, constitutes a quantum wave function collapse. Effectively, the set of all ambiguous constituents of sets is a quantum wave function. Inclusive and/or exclusive distribution of the constituents of the ambiguous subset wave function eliminates it, constituting its ‘collapse’ into unambiguous sets. Quantum wave collapse is logical, then, not ontological.

Distribution of the constituents of the ambiguous set either extends or contracts the membership of the respective sets of which its contents are ambiguous constituents. This extension and contraction is generated respectively by one and zero. One identifies an absolute fullness between two things. Zero identifies an absolute emptiness between two things. One and zero are generated by logical operators, which are states of being. The fullness of one and the emptiness of zero being unobserved, both are imaginative states of consciousness.

An initial state can be extended or unextended. If extended, an infinite number of states are identifiable as extension of it. Since all of these states are so, extension to one is no more necessary than to another. Such a state constitutes entropy. Necessity of extension to one state than to another state is provided by axiomatic limits. All states inside the limits constitute extension, and all states outside the limits do not constitute extension.

If no more than one state is inside the limits, extension is necessary, not probable. This constitutes Newton’s ‘*uniform motion in a right line.*’ If more than one state is inside the limits, extension is probable, not necessary. This constitutes Heisenberg’s ‘probability of transition from this state to another’ (Heisenberg 291).

Range of states inside the limits constitutes a quantum superposition. Extension of state to state inside the limits constitutes a force. Insofar as axiomatic limits are explicitly enumerated, a constitutive mind external to the quantum world is necessary.
Macro Functional Operators

Extension from state to state proceeds by recursion or iteration. A set whose members are identified by similarity to a constant archetype is a recursive set. A set whose members are identified by similarity to inconstant archetypes is an iterative set.

Recursion identifies a set containing no constituent concurrently constituent of another set. Constituents of the set are not concurrently constituents of a subset or subsets identified by differential archetypes. Iteration identifies a set containing constituents concurrently constituent of another set. Constituents of an iterative set are concurrently constituents of a subset or subsets identified by differential archetypes.

Extension from state to state by recursion and/or iteration will compose one of three macro functions. It is by means of these functions that ‘every object [is] linked to its surroundings by indivisible and incompletely controllable quanta.’ Borrowing from mathematics, these functions compose injection, surjection, and bijection.

Injection constitutes a Hilbert space, a substantively and formally unaltered element in different spaces, represented by \( a=a \). Injection is constant content and form. At every point of injective extension, the element will be like at every other point of the extension. Being so, extension of injective element is by recursion.

Surjection constitutes a transmutative space, a substantively altered element in different spaces, represented by \( a=b \). Surjection is at least inconstant content, and at most inconstant form. At every point of surjective extension, the whole will be unlike at every other point of the extension. Being so, extension of surjective whole is by iteration.

Bijection constitutes a transfigurative space, a substantively unaltered and formally altered element in different spaces, represented by \( a_1 = a_2 \). Bijection is constant content and inconstant form. At every point of bijective extension, each element will be like at every other point of the extension, and the whole will be unlike at every other point of the extension. Continuity of bijective part is by injection, and identity of each whole is by surjection. Being so, extension of bijective part is by recursion, and extension of bijective whole is by iteration.

Many Minds

Resolving the quantum paradox with consciousness can be challenged as puzzling. Science is defined by conscious observation by an observer, and verification of conscious observation by an observer by conscious observation of other observers. Conscious observation by an observer is privileged, however. Being privileged, it is unknowable by another observer. Yet, experience is as if content of another consciousness is knowable, and therefore can be verified by conscious observation of other observers.
Although not posing a logical contradiction, posed is a practical inconsistency. Resolution of this inconsistency of observational verification lies in bijection. Presupposed by the conscious observer is observation is shared by other conscious observers, all of whom together compose the community of scientists. Presumed is an a priori axiom of science.

Rendering an infinitude of particulars comprehensible is a finitude of particulars constituted by a limit of the world. Beyond this limit is the observer whereby Bohm’s person observing is identified by disassociation from a limited world. Alternately, a person observed is identified as an autonomous world constituted by a limit within the encompassing limited world.

This limit is identified by the absolute emptiness between two things established by application of the exclusive disjunction of zero. Identifying the person observed is the individual’s limits whereby,

among the attributes an object must have are not only those which it shares with objects of its kind (Aristotelian essentialism), but those which are partially definitive of the special character of the individual and distinguish it from some objects of the same kind (Marcus 191).

Generalized ex post facto, the attributes shared by persons constituent of an encompassing world as a kind composes the self-determined limit of their mutual world.

Content of this mutual world presents a conservation of energy limiting the possible states of constituents of their world. Understood as component of each constituent of the whole, a complete translation of the shared attributes of any one constituent of a world can be made to any other constituent of the world. Distinguishing individual constituents is the potential and kinetic status, as well as sequencing, of the attributes. Now each constituent is distinguishable as a different state of the constant attributes.

Assuming humans fundamentally do not agree is as analytic as assuming they fundamentally do agree. Modeled by this latter is anthropological culture. Deviance can be accommodated by evolutionary biology. Constituent of universal values is a set of tolerated anomalous values, providing a pool of value for adaptation to environmental alteration. Focusing on the aberrant, a stable science is inexplicable. Only by focusing on the common is a stable science explicable.

Bijective observation as defining of membership in the scientific community is for this reason an axiom of critical realist science. Not only is it a requirement of a complete quantum theory, it is foundational to Albert Einstein’s relativity theory. Only by supposing many metaphysical minds linked in observational community is science possible.
Conclusion

Schrödinger’s ‘break with naive realism’ clarifies the observer’s status in quantum theory. Separated from the physical world, the observer no longer suffers the indistinguishability of the quantum paradox. So being, ‘Resolution of the ‘Entanglement’ Result [is] Dependent on the Experimenter’s Intention’ (Schrödinger “11. Resolution of the ‘Entanglement’ Result Dependent on the Experimenter’s Intention.”). Extending the observer’s role, Hugh Everett concludes,

When dealing with a system representing an observer quantum mechanically we . . . . denote . . . by appending the . . . . memory configurations which are in correspondence with the past experience of the observer (Everett 457).

Contracting the observers’ role, because unbound by naive realism, quantum observers share observations.

Despite this is the ‘fundamental requirement of the scientific viewpoint . . . to describe the extra-physical process of the subjective perception as if it were in reality in the physical world.’ Although a pragmatic axiomatic qualification—‘as if it were in reality in the physical world’—this ‘fundamental requirement’ subsequently becomes converted into ‘reality in the physical world.’ Facilitating this evolution is the cognitive scientific reidentification of mind as brain.

Engendered by this reidentification is indistinguishability of the human observer from the rest of nature. Insofar as this is presumed constituent of quantum theory, however, quantum theory is paradoxical. Observation, and analysis and classification of observed, being constituent of consciousness, they are impossible on this assumption, when these activities are intrinsic to science. Descended from Cartesian dualism, science must preserve the mind/body distinction if it is to escape a self-defeating paradox.

Bibliography

Two fundamental assumptions have become dogma in contemporary Anglo-American philosophy of consciousness: that everything about consciousness can be explained in physical terms, and that neuroscience provides the uniquely authoritative methodology for approaching the essential questions. But there has never yet been a successful physical explanation of subjective first-person experience, and reductionism fails to account adequately for thought, reason, and a full range of objects proper to philosophy. Tracking the deep divide within the analytic tradition, I bring a ‘continental’ (German) perspective to bear on recent work from Nagel and Chalmers which shows the reductionist neuroscientific agenda to be incapable of completion, for systematic reasons. Physicalism can explain only structure, function, and mechanism; but self-consciousness is always already embodied and embedded in multiple contexts beyond the structures and functions of brain activity. Consciousness needs to account for itself, to itself, on the terms in which it experiences itself. No explanation of the form provided by ‘neurophilosophy’ is adequate to the most fundamental and essential phenomena of self-consciousness, and neurophilosophy can never philosophically explain or justify itself on its own terms and by its own methodology. These are insuperable limitations for any explanation aspiring to be comprehensive, and such problems have brought contemporary antireductionists in the English-speaking world back around to positions which strongly resemble the ontology and phenomenology of German-language philosophers, particularly Kant, Hegel, and Heidegger.

Two Dogmas of Reductionism

There is a deep division within contemporary analytic (Anglo-American) philosophy of mind and consciousness, between reductionists and antireductionists: those who think there are only physical problems to be
solved, and those who maintain that any physicalist explanation will be fundamentally inadequate to a full account of consciousness and self-consciousness. From the standpoint of German philosophy, I bring a ‘continental’ philosophical perspective to bear on this dispute, aiming to clarify and sharpen recent developments, strongly supporting the antireductionist position, and outlining a critique of what is called ‘neurophilosophy.’

In the philosophy of mind and across the neurosciences, consciousness reductionism in its various forms has been the dominant approach for the past three decades. Consciousness, like everything else, like all there is, is ultimately to be explained in physical terms: ‘reductive materialism is widely assumed to be the only serious possibility.’ The physicalist version of reductionism presents, not arguments, but only ‘hope’: ‘the hope that everything can be accounted for at the most basic level by the physical sciences, extended to include biology’ (Nagel, 2012). In the neuroscientific version of consciousness reductionism, the privileged level of description is neurobiology and the cognitive sciences. Consciousness is to be accounted for as neural activity, which is supposed to afford unique access in answering questions in the philosophy of mind.

Whether in its physicalist or neuroscientific forms, reductionism has a negative connotation, suggesting the dogmatic generalization that everything there is can be reduced to purely material, physical terms—to molecules in motion and nothing more. With respect to consciousness and self-consciousness specifically, reductionism purports to explain these phenomena entirely in terms of the physical processes, structures and functions of the brain: the neuroscientific replacement of the concept of ‘mind’ or ‘self’ by neural activity and nothing more.

From the confident tone of such prominent figures as Dennett and Churchland, one would get the impression that the reductionist agenda has largely succeeded, and that the big questions of consciousness have been essentially answered. Even if all the explanatory details have not yet been worked out, we can be sure that they will be, and we can know in advance what form those answers and explanations will take. The outdated problems, approaches, and language of what used to be called ‘philosophy of mind’ are being replaced by the methodology and corrected terminology of neurophilosophy, the philosophical application of the empirical results of neuroscience.

On the other side of the divide, however, for more than twenty years there have been forceful arguments against reductionism, culminating in the most recent works of Nagel and Chalmers, both from Oxford University Press: a closely argued, nearly 600-page exhaustive account from Chalmers in 2010, *The Character of Consciousness*, and a masterfully brief tour de force on *Mind and Cosmos* from Nagel in 2012, with the provocative subtitle *Why the materialist neo-Darwinian conception of nature is almost certainly false*.

The critique leveled by Chalmers and Nagel shows the attempt to explain consciousness in purely material, physical, or neuroscientific terms to be not
only incomplete but incapable in principle of completion. There are strong, systematic reasons relating to the inherent limitations of physical explanation, why any such account must fail to address what Chalmers (1997) has made famous as ‘the hard problem of consciousness.’ For any assumption to the contrary, the burden of proof is now on the other side; but no compelling counterarguments or rebuttals have been produced, and it seems doubtful whether any could be.

We emphatically do not know that consciousness must be physical. That is neither a law of logic nor an empirical fact. And indeed both sides recognize that the subjective experience of self-consciousness has never yet actually been captured by any physical, materialist reduction. No science has been capable of ‘explaining how and why consciousness arises from physical processes in the brain’ (Chalmers, 1997). Indeed the phenomena most in need of explanation cannot be addressed by any reductionist account, due both to the nature of subjective experience, and to the nature of physicalist, reductionist explanation itself. Despite general acceptance within the sciences and among many scientifically minded philosophers, reductionism is neither self-evident nor unproblematic. For Nagel it is highly doubtful: ‘I realize that such doubts will strike many people as outrageous, but that is because almost everyone in our secular culture has been browbeaten into regarding the reductive research program as sacrosanct, on the ground that anything else would not be science’ (2012).

Neurophilosophy and consciousness reductionism are thus predicated upon two highly problematic assumptions: that everything (including the phenomena of consciousness) can be explained in physical terms; and that neuroscience provides the only reliable and accurate methodology for knowing things about consciousness. Neither of these suppositions has ever been determined, whether by scientific evidence or by philosophical argumentation. Numerous objections stand against these assumptions. Lacking the requisite philosophical justification, they may be identified (following Quine) as ‘two dogmas of reductionism.’

Consciousness Not Explained

Faced with the task of explaining consciousness, as Dennett purports to do, one of the most immediate and important issues obviously concerns the explananda: what exactly needs to be explained and accounted for? This is a fundamental question that directly informs the way the entire explanatory process is undertaken, and how the question is answered will dictate what sort of results are counted as successful explanations. In a work with the ambitious title Consciousness Explained, Dennett (1991) assures us he will ‘explain the various phenomena that compose what we call consciousness, showing how they are all physical effects of the brain’s activities’—but he never does make good on that claim. As he himself must admit:
Adopting materialism does not by itself dissolve the puzzles about consciousness, nor do they fall to any straightforward inferences from brain science. Somehow the brain must be the mind, but unless we can come to see in some detail how this is possible, our materialism will not explain consciousness, but only promise to, some sweet day (1991).

More than two decades later reductionists are still at the point of insisting upon that ‘somehow,’ and still a long way off from that ‘some sweet day.’ A philosophical theory of consciousness cannot simply assert that the only things to be explained are those functions and structures of the brain amenable to neuroscience and physicalist reduction. Reductionism as a philosophical presupposition remains a dogma: it starts from the controversial assumption of materialist metaphysics, with everything that entails. Meanwhile there is overwhelming phenomenological evidence to the contrary that must be explained away in order for reductionism to work. Any adequate account of consciousness must no doubt be informed and supplemented by a thorough understanding of the physical mechanisms of the brain, but a physicalist reduction would definitely be no substitute for a comprehensive philosophical theory, not least because it would fail to address or justify its own fundamental principles and presuppositions.

Ontological and epistemological assumptions about what kinds of things exist and need to be explained cannot themselves be grounded by any appeal to neuroscience: these claims fall within the domain of philosophy proper, and as such they require philosophical argumentation. Chalmers (1997) notes that ‘to establish this position—that there really is nothing else to explain—one might think that extraordinarily strong arguments are needed.’ Yet the proponents of this view have provided no such arguments: rather it has been taken as uncontroversial, even axiomatic. But we cannot know a priori that the philosophical explanation of consciousness will take the form of a physicalist reduction, whether at the neurochemical or quantum level or otherwise, lacking (as we do) the ‘extraordinarily strong arguments’ required for such a high-level determination.

Not only has no such explanation been provided, but there can be none in these physicalist and neuroscientific terms—none that actually accounts for the phenomena in question, rather than sidestepping the real problem and dealing with a different one instead, or denying that there is a problem at all. Reductionism assumes that the only possible objects of knowledge about consciousness are physical and brain-functional, and from this epistemological presupposition draws the further and much stronger ontological conclusion, that what is not explicable in terms of neuroscience and biology, chemistry and physics, somehow does not exist or needs no explanation. That is not a scientific truth, nor could it be established by any science, not even in principle. The claim that scientifically derived empirical facts are all that we
can know about consciousness is not a scientifically derived empirical fact. As Chalmers (1997) responds to his critics:

*Proponents of the ‘no problem’ view sometimes like to suggest that their view is supported by the results of modern science, but all the science that I know is quite neutral here: I have never seen any experimental result that implies that functions are all that need to be explained. Rather, this view seems to be rooted in a philosophical claim...*

The subject of first-person mental life—as experienced prephilosophically, given prior to any analysis—must be central among the default explananda in the philosophy of consciousness, until and unless it is proven (as it never yet has been) that the subjective selfhood we experience is in fact an illusion. Otherwise that remains an extremely counterintuitive (perhaps even incoherent) position, directly contradicted by immediate evidence, to which one is driven only because the assumptions of physicalism require such a denial. For the antireductionists, far more plausibly and practicably, ‘conscious subjects and their mental lives are inescapable components of reality not describable by the physical sciences.’ These realities cannot be avoided or dismissed by any account of consciousness that aspires to be anything close to comprehensive: conscious minds ‘are among the data that a theory of the world and our place in it has to explain’ (Nagel, 2012).

**Accounting for the Nonphysical**

It has long been argued in at least one main line of Western philosophy—and in numerous nonwestern traditions—that not everything which has reality and significance is physical. From Pythagoras and Plato to Descartes and Kant, the rationalist tradition has always had powerful arguments for the existence and reality of the nonphysical. Still, the intractable problems of mind-body interaction have led empirically oriented philosophers to try to avoid dualism at all costs (Dennett, 1991). But it does come at a cost: anything apparently not physical must be *epiphenomenal*, essentially denied reality. This encompasses all evidently nonphysical features of self-consciousness, subjective experience, thought, rationality, language, ‘the mind’ itself. It is no coincidence that Plato and Descartes are among the least reputable in contemporary neurophilosophy. But reductionists remain confronted with obvious problems of ontological plurality, even if dualistic minds do not exist but only brains. Reductionism is already incapable of accounting for the indisputable actuality of mathematics and logic: both are clearly related to the physical, but also retain their formal relations and truth apart from and independent of particular concrete physical expressions or examples. Math and logic have objects which appear to be distinctly nonphysical, nonmaterial, and to have a pronounced universal
character: they go beyond the physical, and thereby call into question the explanatory range and power of any physicalist reduction. There is an unreconciled dualism, *a priori*, at the heart of any supposed materialist monism.

Reason itself is another central philosophical concern, related to math and logic, which does not appear to be physically reducible, but which is obviously operative and capable of generating truth and knowledge of all kinds, some ‘pure’ (in Kant’s sense) and some ‘practical.’ Reason must be accounted for on its own terms, for only with reason can we explain, correct, or answer anything. Reason is its own ‘final court of appeal’; it has ‘completely general validity, rather than merely local utility’ (Nagel, 2012). For ‘practical’ purposes we must be realists about reason: the rational is actual, and rational relations obtain independently of any brain activity. Reason is one of a whole broad set of realities that we experience and need to explain: Nagel (2012) includes ‘consciousness, intentionality, meaning, purpose, thought, and value’ along with reason, as chief among the explananda incapable of physicalist reduction, for which a higher-order level of ontological description is required.

If we add to that list other essential philosophical concerns like language, art, religion, literature, history, culture, ethics, politics, justice—it becomes more and more apparent that no scientific reduction is going to be able to address and explain all these, and most certainly not a neuroscientific reduction. But all these areas of inquiry are the rightful domain of philosophy: they are part of the broad set of interconnected questions that define philosophy, and they show just how limited neurophilosophy must be considered in comparison with philosophy proper.

The empiricists have never been able to account for conceptual reality empirically, not in any convincing way, because pure forms of reasoning and inference are inherently abstract and *a priori*. The intelligible realities of mathematics and logic cannot be denied actuality, yet whatever ‘being’ they have must be other than physical. They may have physical *instantiations*, but what is thereby instantiated is *conceptual, universal*, and any adequate ontology must be able to explain the fundamental reality of reason, logic, abstract concepts, universals. The very existence of logical necessity and the fact that true propositions can be deduced from other propositions is itself something that must be accounted for. Any and all empirical arguments must presuppose *a priori* principles of reason, whose justification is not biological, and which are not reducible to the terms of physics or any natural science. If reason, math, and logic are nonphysical, then not everything is physical—so then not everything can be explained in physical terms.

Math and logic present confounding ontological counterexamples, if not a flat-out refutation of physicalist, materialist monism. They reveal the insuperable limitations of reductionist explanation: dualism is already an inescapable problem, before the even harder problem of subjective experience and self-consciousness is factored in. If physical reductionism cannot account ontologically for the abstract, nonphysical and universal truths and objects of
math and logic, *a fortiori* it cannot explain the forms of rational consciousness which apprehend those truths and objects. So now there are at least three fundamental ontological categories to account for: (1) the conceptual (e.g. mathematics), no less than (2) the material/physical, and (3) self-consciousness. It is striking how directly these categories recall the Hegelian triad of Logic, Nature, and Mind [*Geist*]: we appear to have circled back around at a deep level to German philosophy, as neo-Kantians, neo-Hegelians and neo-Heideggerians.

Consciousness, thought, reason, math and logic join what is a long list of ontologically significant phenomena whose being cannot be explained or accounted for in purely physicalist or materialist terms. Not only mathematical and logical entities are nonphysical; so too are any abstract concepts, including those of ethics and value: persons, rights, equality, dignity, fairness, justice, ‘the good, the true, and the beautiful’—these all go beyond the merely physical substrate. This is not to say that they must be ‘supernatural,’ in the derogatory or reproachful sense of the word. It means that they cannot be explained in physical or neuroscientific terms: an expanded set of ontological categories and an expanded conception of ‘nature’ are necessary to overcome the explanatory gap. To adequately account for the phenomena of self-consciousness, *phenomenology* is necessary: as Chalmers (1997) argues, phenomenological approaches to the ‘hard problem’ must be ‘absolutely central to an adequate science of consciousness: after all, it is our own phenomenology that provides the data to be explained!’ Any Anglo-American attempt at explaining consciousness must reckon with the critique already advanced by the German philosophical tradition, concerning the foundations and general features of the phenomenological approach: not only in Heidegger’s or in Husserl’s sense, but primarily along the lines of *Hegel*, that original post-Kantian phenomenologist of consciousness.

**The Impossibility of Neurophilosophy**

The neurophilosophical project was therefore bound to fall short, because neuroscience cannot substitute for phenomenology or ontology: it cannot address or explain entire areas of philosophical inquiry. No science has ever scientifically established its own philosophical presuppositions: the sciences do not empirically justify the logical, ontological, and epistemological determinations which are the conditions for the possibility of science.

Of course the philosophy of nature has branched off over the centuries into many sciences; but science has never been able to ‘replace’ philosophy, and there is no reason to expect that neuroscience will fare any better. Philosophy and science have a great deal in common, and there is a long and likely familiar set of arguments for their similarity and affinity that need not be rehearsed here. But there are also numerous points differentiating philosophy from science, which show how and why they are not the same thing; science on its
own cannot establish philosophical conclusions any more than philosophy on
its own can establish scientific propositions. But philosophy, crucially—as
Kant and Hegel practice it, for example—can *account for itself* on its own
terms, while the sciences cannot.

From the overarching view pejoratively termed ‘scientism’—so called
because it dogmatically insists that the sciences can explain everything—we
can coin the term *neuroscientism*, to designate the dogmatic extension of
neuroscientific reductionism into other domains of philosophy, the attempt to
appropriate and reframe the ‘traditional’ problems of philosophy into questions
that can be addressed by neuroscience. *Neurophilosophy* is meant to be a
*science*, as Churchland (1989) makes clear in her subtitle: *Toward a unified
science of the mind/brain.* More recently, Churchland (2008) claims that
traditional philosophy of mind has been taken over and ‘replaced’ by
neuroscience:

*Since the weight of evidence indicates that mental processes actually
are processes of the brain, Descartes’ problem has disappeared. The
classical mind/body problem has been replaced with a range of
questions: what brain mechanisms explain learning, decision
making, self-deception, and so on. The replacement for ‘the mind-
body problem’ is not a single problem; it is the vast research
program of cognitive neuroscience.*

But ‘the really hard problem of consciousness,’ according to Chalmers
(1997), ‘is the problem of *experience,*’ and it has assuredly not ‘disappeared’:
‘any neurobiological or cognitive account will be *incomplete*, so something
more is needed for a solution to the hard problem.’ For Nagel (2012) ‘the
mind-body problem is not just a local problem’: far from having disappeared,
‘it invades our understanding of the entire cosmos and its history.’ The
implications are profound, and the explanatory task that much greater, when
the ‘really hard problem’ takes us from neurons out to ‘the entire cosmos and
its history.’

Few neuroscientists would presume to make pronouncements concerning
the cosmos. Neurophilosophy is not in a position to give a comprehensive
analysis of consciousness in its cosmological context, considering its full
implications. Even the object directly at hand remains a ‘really hard problem,’
an *unsolved* problem: showing how and why the mind is nothing more than
brain activity. But the ‘impossibility of neurophilosophy’ refers more broadly
to the inability of cognitive neuroscience to give an adequate account of
fundamental philosophical issues. It does not and could not address entire
branches of foundational questions on its own terms. Neuroscience cannot
provide a systematic and coherent integration. As a philosophical methodology
it is inapplicable and incapable outside its area; it certainly cannot presume to
provide an account of the totality. In these respects neurophilosophy is and
must remain limited and incomplete, due to inherent differences between science and philosophy proper.

Scientists have many times tried to claim epistemic priority for science over philosophy, but such claims are themselves philosophical and therefore self-undermining. Science has always been closely related to philosophy, but it has never been adequate to ‘replace’ it, for systematic reasons relating to the kinds of empirical questions settled by the sciences: causal questions, physical, material, functional questions. Their answers are framed as scientific explanations or hypotheses. But those are not the forms of fundamental philosophical questions and answers, a difference which is essential to both philosophy and science. The normative claims of philosophy (including the priority given to scientific knowledge) are not like the empirical propositions of the sciences themselves, and they are established in very different ways. The broad categories of philosophy and science are as obviously distinct as they are also in subtler ways similar. A scientific fact discovered e.g. by functional magnetic resonance imaging (fMRI) does not answer a philosophical question; and conversely if it is a philosophical question, then the answer is not discoverable by laboratory experiments or fMRI scans. Among other things, philosophy asks metaquestions: metaethical or metalogical or metaphysical questions are not settled by fMRI or by describing how brain mechanisms work.

Neurophilosophy thus represents a branching off from some areas of inquiry in philosophy, while leaving other entire areas completely open and untouched. It provides essentially no account, for example, of logic or ontology, and no ethics or ethical theory that is not already question-begging. Its restricted range offers explanations for only a subset of issues within the philosophy of consciousness. The entire orientation and approach appears excessively ‘neurocentric’: overly privileging the brain, yielding an ontologically incomplete set of phenomena as compared to the full range of objects of philosophical investigation. It is evident that fundamental questions, metaquestions which are properly philosophical, could never be settled by an appeal to neuroscience or indeed any science. There is no scientific method to ground, validate, or verify metaethical determinations: rather such evaluative justification must be borrowed from this or that philosophical theory.

Philosophy, on the other hand, encompasses activities in multiple contexts, and deals with objects and concepts for which multiple levels of description are necessary beyond the physical, material, or neurobiological. Philosophy requires reasoning, argumentation, evaluation, analysis and synthesis, interpretation, hermeneutics: it involves broad concepts like meaning, value, significance. It is concerned with social realities and abstract systems no less than the biological, chemical and physical substrates of such higher-order structures. Philosophical accounts have to make sense of phenomena which are social, linguistic, cultural, political, experiential, aesthetic, evaluative, normative: these are generally not delineable and determinable the way the objects of the sciences are, and they are not explicable in the same terms.
Scientific data may be relevant to such analyses; but they can never be decisive or dispositive: philosophy is larger than science, and comprehends it (in both senses of the word). The objects of philosophy include higher levels and larger contexts, the social structures, institutions, organized practices and collective functions—including philosophy itself—which happen not just in brains but in ‘the external world,’ with everything that entails.

In Heidegger’s terms, Dasein or ‘human being’ is always already being-in-the-world (in-der-Welt-sein), the phenomenology of which requires philosophical investigation and explanation: a full accounting and reflective interpretation which provides for the ‘lifeworld’ within which human being takes place, the multiple contexts larger than the given physical immediacies of brain activity. Philosophy can, where neurophilosophy cannot, address the reality of complex, externalized, formal systems, structures, and narratives, transcending individuals and encompassing not just ‘many brains’—the fallacy of composition—but going beyond brain activity altogether. Oddly, Churchland (2008) seems to agree to this when he writes: ‘Solving social problems is an awesomely complex business, requiring relevant facts, including facts about cultural practices, about what brains do value, and fact-based predictions about consequences. Fundamentally, moral/social problems are constraint-satisfaction problems at the many-brain level…”

**Back to Kant, Hegel, and Heidegger**

I have taken note of some clear parallels between contemporary antireductionist arguments and important positions maintained by Kant, Hegel, and Heidegger. A more extensive treatment of this topic would bring out the many detailed points of similarity connecting Nagel and Chalmers with their predecessors in the German-language philosophical tradition, who also include (among others) Leibniz, Schelling, Nietzsche, Husserl, and in his own way, Wittgenstein.

Here at least a few of these parallels may be briefly drawn. In Hegel’s analysis of explaining consciousness, for example—what may be his most essential philosophical project—a diremptive yet self-reflexive relation obtains, between the conscious object which is to be explained, and the conscious subject which does the explaining: the same thing accounting for itself to itself. We have to explain ourselves, as the knowing, experiencing beings who are capable of explaining and accounting. Making intelligible one’s own being and knowledge of oneself must be part of one’s general account of the world. We are the being that does the explaining of being, including our own: we are the irreducible consciousness that has come to understand itself as irreducible consciousness. Knowing ourselves for what we are, the self-conscious subjects as well as the objects of inquiry and knowledge, clearly suggests Hegel’s Mind or Spirit, coming to know itself as Mind or Spirit—as ‘Substance just as much
Two Dogmas of Reductionism: On the Irreducibility of Self-Consciousness and the Impossibility of Neurophilosophy

as Subject.’ Nagel (2012) acknowledges his debt to Hegel and German philosophy, sounding especially Hegelian when he writes (for example):

Mind ... is doubly related to the natural order. Nature is such as to give rise to conscious beings with minds; and it is such as to be comprehensible to such beings. Ultimately, therefore, such beings should be comprehensible to themselves.

The subjectivity implicit in self-consciousness and self-comprehension requires the reality of reflexively self-aware persons, conscious selves, who have first-person experience. Following Kant we could even call these the a priori conditions for the possibility of explaining consciousness. Not only is the subject contained a priori in any and all such explanations; so too is the self-conscious rational community of inquirers, for which truth, knowledge, scientific explanation itself are all real values. Only on such an assumption does the entire explanatory enterprise make any sense. These are among the phenomena of consciousness that need to be explained, to which no reduction to brain activity could possibly be adequate.

German-language philosophers had long recognized and argued that empirical research into psychology, anthropology, biology and the natural sciences (necessary as it is) would never—could never—be sufficient for a philosophically adequate account and comprehensive explanation of ‘mind,’ self-consciousness, thought, and reason. But philosophy in the English-speaking world had to come around to these conclusions independently, that the subject of first-person experience is ‘transcendental’ in Kant’s sense—that self-consciousness is ontologically irreducible, that it is no epiphenomenon but fundamental to the nature of the universe. The results, as Chalmers (2010) writes, are ‘liberating’:

This I take to be precisely the liberating force of taking consciousness as fundamental. We no longer need to bash our heads against the wall trying to reduce consciousness to something it is not. Instead we can engage in the search for a constructive explanatory theory ... which accommodates consciousness in the natural world. And a fundamental theory of consciousness ... is the best way to do just that.

Assigning consciousness its proper place as fundamental, the consequences are profound: we have to expand our ontology and become realists about a broad range of possible objects, not all of which are physical. This will require a substantial revision to our entire conception of the universe: if consciousness is real, then materialism as the whole truth is ‘almost certainly false.’ Just as with Kant, Hegel, and Heidegger, reason, ontology, phenomenology, and social theory are all advanced again as central and necessary, as ineliminable in the project of explaining and accounting
adequately for consciousness. These long-held ‘continental’ positions may now be said to have been vindicated from within the analytic philosophical tradition itself.

Bibliography


For a long time, with several peaks determining history, recognition and the essential capabilities of the human mind were linked to the assumption (and resulting dichotomy), that the performances of sensation were pure or self-effective. Those of the intellect however, were understood to provide the counter-side of a similar condition related to a proper and contrary source. This dichotomy has seen a revival in connection with the writings of McDowell and the question of non-cognitional versus cognitional content. It is also more or less closely related to the philosophy of the mature, critical Kant.

The proposal for amendment follows three aspects: (i) the definite separation of the faculties from their performances; (ii) the issue of the adhibition of one faculty upon another according to Kant - instead of the rational one-way application (in connection with the categories), the solution depends upon the convertibility of the faculty relation, i.e. sensation is also able to determinate the mind; (iii) the relation between sensation and the intellect is seen as real differentiation (or the outcome of a polar relationship) which, by reason of necessary limitation, contains some narrowing of McDowell's position.

1

It is pure experience that tells us, before any reflection occurs, that objects are divided up into certain groups and have certain qualities. It is consciousness that gives us the immediate value of objects, and there are ‘empirical values of unreflective life.’ The opposite is the case: truth is the product of reflective thought which, as is known, only occurs when the system of values is disturbed. At this level, it is no longer the very nature of the world which dictates our conduct, but reflective thought (Durkheim 1983, 55).

Pure existence as the differential limit of reflected thought is not a transcendental claim. Depending upon the differentiation, it is also not completely within pragmatism, which takes pure experience as a parallel and real condition of access. Otherwise, pure intuition belongs cardinally to the philosophy of Kant. In the first place, it means imagination is capable of being related to mathematical thought and geometrical as well as empirical sensation.
So the issue is divisive because consciousness and recognition are bound to a counter step of sources as being their correct level of foundation. In addition, it is no wonder that the philosophy of Kant is well suited to incorporate novel branches of interpretation. This seems significant evidence (at least latent) that contemporary philosophy (on the whole) is still grappling with a field, a margin (and mountain-like area as well), which Kant has delineated and bequeathed. Where mountains - and extraordinary heights in particular - are waiting, an abyss might well be encountered. The visitor required for its contemplation (which depends upon the point of view), is thus far the consequential as opposed to the principal issue, which points at once to the metaphysical obligation: the non-accessibility of the real - the so-called ›thing in itself‹ - and its proper composition and constitution.

This other problem - the abyss or ‘hiatus’ as Fichte has called it in his later self-interpretation of the Wissenschaftslehre - is fueling skepticism on a second level of mind (so to speak), to the extent that Kant himself - with his critical attitude - has founded the first level concerning the natural and inherited means of human reason. It still motivates a couple of novel theories and solutions in the approximate background, whereby it is not necessary for the theory concerned to announce itself as being an interpretation of Kant. They are tackling a problem - properly the reconciliation of sensation with the mind. This should be acknowledged as the core issue of Modern Age philosophy and the entire endeavour of science. The differential problem is also closely related to the introductory consequential issue - the metaphysical obligation. Normally, the standard view is that the senses are responsible for the accountability of reality, whereas the mind is responsible for the counterpart - the so-called ›space of reasons‹ (Sellars vs. McDowell). This term has become rather famous in connection with a theory of segmentation, fragmentation and finally of a splitting that has been termed the opposition between non-conceptual and conceptual [re]cognition. Without any acknowledgement that this opposition might belong to the polar one, several authors around John McDowell - who is the main target and leading figure - take a position for or against the possibility of conceiving the world in terms of intuition, i.e. without the assistance or salient influence of concepts (and of the categories in particular) (see McDowell 1994 and 2009). The target - conceptualism - is also the medium of recent interpretation around skepticism and coherentism. This is no wonder, because both share the issue of the reliability of perception as the most comfortable source of true beliefs, as opposed to otherwise coming to terms with the intricate problem of reconciling the transition from sensation to the mind. McDowell himself has succinctly founded his position not only on the precincts, but rather a core area of critical philosophy - the relationship between concepts (Begriffe) and intuition (Anschauung) (McDowell 1994, 41-45). So, one has good reason to test and/or rest with this standpoint because - according to one’s perspective - it predetermines or overshadows the others.
Amendments to the Theory of Recognition

Sensation is already concept-laden. This is the orientation McDowell is pursuing and there can be no doubt that his eclectic methodical reasoning is responsible for a large amount of critical reception. Settling mainly with Kant, one can also discern a primordial introduction to Wittgenstein and an explanation of his standpoint according to Hegel. Because one cannot understand these affiliations without a clear description of his main approach and goal, this shall have first place. The description that follows has two parts - one concerns the terminology, whilst the other concerns the arguments. There is also a region in-between which provides the main orientation incorporated by his dispute. The terms one should have a close look at are ‘spontaneity’, ‘receptivity’, ‘thought’, ‘intuition’, ‘sensation’ and ‘experience’. There is a second group stemming from contemporary philosophy (Sellars in particular) involving the ‘space of reasons’, the ‘space of concepts’, ‘conceptual content’ (which implies a direct negative), the ‘non-conceptual content’ and ‘the Myth of the Given’ in particular. Judging from historical background, this last concept may signify the manner of dealing with mature Kantian philosophy and its main conclusions. It must represent both the thing in itself and the overall non-accessibility of the real, where its elements are located as soliciting the impressions of the mind or “given intuitions”. Therefore, one could not assume that the whole sphere should be termed a myth. In terms of Kant it is not. It is essentially the contrary (otherwise the impressions would lose their real ground). The ‘Schein’ (appearance) - one of Kant’s most well-elaborated concepts - would merge into ‘Irrtum’, ‘Unwissenheit’, ‘Unvollkommenheit’ (error, uncertainty, non-perfection), the totally imaginative (necessarily unreal) - if held as equivalent to myth. Accordingly, the last two centuries have seen a large amount of interest in this cardinal outcome of critical philosophy and the accessibility, reliability and even acuteness of the transcendent realm. This also includes natural science and mathematics, Helmholtz, Weyl, Frege and Klein, due to new conceptions of intuition and/or their properties relating to the continuum. Viewed from this perspective, Kant’s classical position concerning the function and settlement of experience was never altered into a non-classical or even a non-critical one, as myth would have to have been classified with utmost justification.

Now, before getting to grips with the terms and the arguments, it would be helpful to cite samples from the aforementioned inter-region. These are: (i) ‘judgments of experience grounded in a way that relates them to a reality external to thought’ (5) or ‘the bearing of empirical judgments on reality’ (5); (ii) ‘the external constraint on our freedom to deploy our empirical concepts’ (6); (iii) ‘empirical substance infused into concepts’ (7); and (iv) ‘impingements by’ or ‘impacts of the world on our sensibility’ (10-11), both close to ‘impingements on spontaneity by the so-called conceptual deliverances of sensibility’ (13). Finally, there is (v) the interpolation of ‘rational constraint’ versus ‘only causal influence from outside’ (14, 17) and one should include the
‘fear [that thinking is getting] out of touch with the world outside us’ (McDowell 1994, 17).

Beginning with this last strand of concepts, it should be evident that McDowell is looking for a perspective that enables the philosopher (and anyone else with a concise reliance upon their own experience) to occupy a position which does not equate to realism. In several writings, McDowell has followed the lines of empiricism with regard to the intrinsic impact of the philosophy of mind. One should understand that, even if he sometimes indulges in a monistic Hegelianism, he is asserting the impact of reality on the mind. In other words, that any valuable theory concerning philosophy has to guarantee (in practice) the impact of reality on the ‘space of concepts’ or - to put it once more in his own terms - ‘to acknowledge an external constraint on the exercise of spontaneity in empirical thinking’ (McDowell 1994, 50-51). His following solution is not very comfortable because he does not immediately subscribe to dualism, as Kant did. Instead, he teaches that (i) the ‘space of concepts’ is not smaller than the ‘space of reasons;’ and (ii) that the extension of this conceptual sphere is - in philosophical terms - non-negated or never negative: it stretches to the very limits of experience where representations are formed. He therefore sees himself as justified to conclude that sensation is always concept-laden, or ‘in experience one finds oneself saddled with content' (McDowell 1994, 10). This solution is in accordance with his later paper ‘The Logical Form of an Intuition,’ where he holds that, faced with the possibility of ‘sheer sensation’ (which Sellars demands according to the sources of Kant), there must be a peculiar ‘togetherness’ of conceptual capabilities within an ostensive act of judgment. ‘An ostensive seeing [!] that there is a red cube in front of one would be an actualization of the same conceptual capacities that would be exercised in judging that there is a red cube in front of one, with the same togetherness' (McDowell 2009, 30-31; emphasis added). This togetherness has to be kept in mind for the amendment because it is also stressed as a ‘counterpart to the logical togetherness of the “red” and “cube” in the linguistic expression of the judgment - “There is a red cube in front of me”’ (McDowell 2009, 30). What has to be separated from this issue of the logical togetherness of the sensual versus conceptual capabilities, is the question of sheer sensation, whose limiting factor is also denied by McDowell. Nevertheless, these ‘same’ capabilities are still the main problem, to the extent that circularity has to be avoided. So far, the togetherness seems only an analogical clue on descriptive grounds or a simple reminder of Kant’s formulation of transcendental analytics (from which would follow that this ‘sameness’ is the result of categorical impact or the impact of the pure categories). Finally, one should also keep in mind that the limitation is not properly expressed as a token of differentiation. Even more so with polarity, which otherwise (especially in accordance with the ‘togetherness’) seems very fit to capture the opposition of parallel and necessarily bound capabilities. This holds especially true when they are exhibited in a context of overt focalisation (ostension). The question of negation remains bound to its real source and the
Amendments to the Theory of Recognition

stemma of objectivisation or the ‘ungleichartigen Bestimmungsgründe’ (Kant. Logik Dohna-Wundlacken. In AA XXIV, 721).

3

After this first explanation it is possible to address the terminology. His manner of writing is such that he normally does not cite the historical issues, especially concerning Kant. In order to be able to follow his argumentation, one must therefore contrive or immediately subscribe to the meaning of the terms he provides in the context of the mature, critical Kant. Of the highest importance are his uses of “spontaneity” and “receptivity”. As should have become clear, he does not define them. Spontaneity, however, serves as the main instance of rationality, to the extent that it represents the ability of human beings to deploy their conceptual capabilities. It is the pivotal axis where consciousness is able to make a classified and contrary decision to either follow its perceptions and non-reflected thoughts, or to deploy its conceptual means: the first is prevailingy receptive, the second a spontaneous consciousness which is largely along the lines of Kant and tradition. Nevertheless, McDowell avoids any formulation or conception throughout, where concepts or conceptual means are properly applied to perceptions and sensual deliveries. This should be the first place to look for amendment, not only because of the historical background (where perception and apperception according to Kant - include the adhibition of pure and/or empirical concepts as instances of the faculty of mind to sensation), but also due to the possibility that differentiation between faculties is as real and therefore theoretically required as the aforementioned ‘interpolation of the rational constraint.’

According to this pivotal axis or ‘faculty of spontaneity,’ (McDowell’s ubiquitous composite term) one has to presuppose that a human being - especially when being erudite or well-suited to engage this faculty - is able (i) to achieve the very limit of consciousness where concepts are necessary in order to register any experience; and (ii) is without proper constraint (or limitation) free to do this: that is, it is spontaneity in particular, which guarantees that consciousness always has the specific power to insert and activate its conceptual capabilities. Concerning the amendment in the next part, one should re-assure that both sensation and the intellect are genuine faculties, but also that McDowell constantly calls spontaneity a faculty. Because this has another ranking in relation to the other faculties or capabilities, the whole matter should be transferred to logic. In this light, McDowell does not investigate (i) into transcendental versus formal logic, (ii) the commitment of imagination or Einbildungskraft in connection with consciousness-forming schemes or schematic content, and in particular not (iii) into the constant use or impact of the categories (in ‘The logical Form of Intuition’ he calls them a difficult matter, at least something that must be separable from the main issue of how to deal with immediate perception and if this bears a conceptual impact or not) (McDowell 2009, 30). Even if one might hold that they are either
obsolete due to historical development (especially in the analytical philosophy and logic since Frege), or not exempt from internal/external fusion (hence finite ambiguity concerning other concepts or the real stemma), this abstinence from proper investigation into the categories seems (on the whole) inconsistent. Taking into account his conviction that experience, even at the very limit of sensual input or interference, is bearing conceptual content, one should investigate into the possibility - or (according to Kant) the necessity - that this bearing has to do with categorical impact from the beginning, i.e. in each and every case. In other words (to use Kant’s teaching) pure concepts have to be applied to perceptions and sensual deliveries in order to become recognition. On the contrary however, McDowell always concentrates - in discussion with and in counterpoint to his colleagues’ positions - on the realm of empirical concepts. No doubt, this will make a difference in the case of animals and fictitious beings, as these are not able to provide evidence for identifying spontaneity as a so-called ‘faculty of conceptual understanding.’ Furthermore, McDowell uses receptivity in parallel to sensation as the passive foundation of consciousness. His construction is such that experience as a passive power invokes sensations but that this passive power has already equalised. Due to an assigned or borrowed transcendental necessity (so to speak), it is steadily assimilating conceptual capabilities. Even if he does not speak of a transcendentals necessity, this is the reason why he does not acknowledge a margin (or a differentially external realm of mere sensations or pure emotions) comparable to the other axial direction, according to which (like Sellars) he does not acknowledge a realm of pure concepts.

4

Conclusion. Part 1. To come to a conclusion, the categorical issue is not the main difficulty that his conception implies. The achievement of the utmost level and eventual limit of consciousness - where concepts might show their genuine stemma within a contrary and broken, interrupted or continuous network of other concepts - is opposite to the other limit, which is his constant focus: the most concrete, the very sensual input where also the expressive powers of language are challenged. Nevertheless, several times the main issue is connected with logical problems. The first one is that his solution does not provide the means to have full insight into the orientation, the direction or the proper issue of perspective: insofar as one presupposes the passive power or constraint from external ground(s), he simply and permanently equates it with conceptual impact. This, however, represents the contrary orientation because it is driven and powered by the internal (re)source, the power or ‘faculty of spontaneity’. Hence - to be fully rational or completely understandable - one must know how the contradiction (necessarily external versus necessarily internal) is really resolved. In any instance where a judgment like ‘this red cube
Amendments to the Theory of Recognition

in front of one’ is formed, both directions must coincide and even fuse within a shared focus. This fusion must be systematically set apart from any confusion. It is therefore better to speak of a polar relationship and pursuant opposition overall, instead of a contradiction, as McDowell also has constant access to differential language. Secondly, because he does not systematically distinguish any logical order within spontaneity - each of the faculties, further their performances or renderings, i.e. conceptual items in relation to items of sensation and/or perception - one should for a moment return to the conception of Kant. Contrary to McDowell, Kant always postulates that concepts are adhibited or applied to intuitions or phenomena belonging to the sensational realm (see Critique of Pure Reason. In: AA IV, 100 ). Consequently, when isolating or properly differentiating spontaneity from receptivity - both as overruling determinations of human consciousness with regard to their peculiar performances - one should reserve the overall possibility

that sensation or experience (as the realm where initial perceptions come into the mind), is determining the counterpart - the conceptual capabilities.

Accordingly, determination exists on a par with spontaneity. However, spontaneity is not restricted to being solely activated, i.e. engaged in connection with conceptual capabilities. Quite on the contrary, (and complying with the fact that the relation between sensation and the intellect is due to a polar opposition), human consciousness is spontaneously distinctive when it feels or perceives something and applies its impact on the mind or conceptual capabilities in order (i) to select, (ii) to purify and inform or (iii) to simply find the correct concept (according to the token or base of memory): in other words, to determinate the intellect by sensation.

This is, of course, also contrary to the decision that Kant made. He only taught that by reason of the subordination of intuition or phenomena to the mind via application of the categories, a determinate recognition will follow (see Critique of Pure Reason. In: AA IV, 99-100). But he did not fully investigate (or otherwise require) the immediate, converse counter-direction: that sensual deliveries determine the mind, so that consciousness (in this case) follows the other (polar) orientation. One main obstacle to this approach was that he constantly took sensibility on esthetic grounds, so that logically it is restricted to coordinate aggregates without a proper grip on subordination as the counterpart. Nevertheless, pursuant to the logic of the alternative, human consciousness (with regard to recognition) is also principally clear and distinct in its orientation. It is not forced to first acknowledge a necessary input from outside and to immediately fuse or equate it with the counter-direction, postulating that the same impact is equivalent to conceptual content or capabilities, respectively. Quite on the contrary, the ordering of the faculties has to be separated from any ordering of their performances - i.e. concepts, proper imaginations and sensations - so that determination steadily builds two main fields of manifestation: their transcendental (or necessary) peripheries (or
‘sphere’) according to Logic, §12, where Kant teaches that ‘concepts belonging
to the same sphere are called reciprocal concepts’ (Wechselbegriffe; conceptus reciproci (see Logic, § 12. In: AA IX, 98). Due to ambiguity, it seems necessary
to make the distinction from polarity. Thus the common sphere may be the
periphery of each and of course it is not necessary to separate it from the
possibility of joining sensation and intellect). Presupposing the determinative
impact - and when it is correct that the mind implies (or ‘subordinates’
according to Kant) intuition as a faculty - then it is equally correct to say that
intuition is able to imply (or subordinate) conceptual capabilities before any of
these capabilities are instantiated within a concrete realization: this should be
the principal distinctive ordering of consciousness. In this way, determination
has a general settlement or seat within the faculties and their relations and
spontaneity (on the highest level) is not prevented from making the mind
(uncrooked) subordinate to sensation, in particular in connection with any
novel formation of unused concepts or in any instance of memory associated
with proper re-cognition, a field exhibited by most of McDowell's examples.

Part 2, consequences and demonstrability. “If my recognition has to be
related to the object, then sensibility and the intellect unify. The one provides
intuitions, the other concepts. Now, if two forces are acting; then therefrom the
third one will result. The sensibility is flowing in with any act of the mind, and
hence a certain direction of the mind will herefrom originate. Thus sensibility
is the cause by reason of which faulty recognitions will be taken for true” (Kant, Logik Hechsel. In AA, vol. XXIV, author’s translation). This citation
from one of the transcripts of Kant’s logic may serve to introduce two main
aspects. First he teaches that sensibility and the intellect are always connected
or interact with one another. If one takes the application of forces (instead of
faculties) for real, then also the polarity opposition would be correct
differentiation included). Otherwise, the last sentence has to be subjected to
critical scrutiny and the alternate theorem that sensibility is capable of
determining the mind within the complementary and fundamental relationship
of human consciousness. Indeed, the resulting force or faculty relation should
include the two main determinations as strictly possible,

i. mind ⇒ sensibility (including imagination) (the ‘normal’)
ii. sensibility (including imagination) ⇒ mind (the spontaneously
convertible);

and consciousness may then primarily rely upon the one or the other.
Nevertheless, when sensibility determines the mind, a large field of application
must follow which moves beyond the sphere of imaginations and mathematical
intuitions (with regard to Kant or primarily Euclidian geometry). In addition,
the issue of falseness has to be resolved from involved determination and their
possible conversion and congruence which, as the convertibility of
consciousness according to its faculty relation, should also represent a great
advantage of this approach or theorem (see Wilkens 2002 and forthcoming).
Amendments to the Theory of Recognition

When the mind is deceived by the senses (or these invoke an appearance) then the determinate relation or inclusion of the mind is not as correct as it should be: curved (non-straight), confused (in this peculiar sense, partially merged), broken, shifted or otherwise extra-polarized. Consciousness is normally free to create and inform both forms of determination in order to calculate the congruence of this main parallelism (which is due to the convertible relationship).

The field of application is plurifold - to avoid manifold intuition according to Kant - which should remain a heuristic concept:

(i) Sensibility - insofar as it is coupled with spontaneity and determinant power - must be capable of classifying. Otherwise, there must be sensual relationships which include their proper subordination. Sensibility does not represent (and is not restricted to) a necessary accidental aggregate due to essence. Kant, in addition to the issue of mere ‘coordination’ (taken esthetically as non-logical), did also not investigate the possible merging (lawful fusion) of representations. Hence the logical root of consciousness is located within the centre of the faculties and not only within linguistic capability. The human *logos* over-spans the faculties instead of settling only with the mind.

(ii) The evidence and/or proof of sensual classification can be found in a large field of empirical applications already beginning at the time of Kant and expanding ever more with the industrial (r)evolution: plans and drawings (in the exact sense), hence coupled with architecture, machinery, circuits, topology (land surveying), cartography and process diagrams in the exact sense of today (BPMN). Here the exact sense is synonymous with the instruction of how to understand real or objective relationships. Pursuant to architecture, p.e., the plans have legal force or priority over the textual description of the same facts. Whether this legal priority is included in the contracts or not, an engineer knows (due to the plans and drawings) how a machine or building effectively has to be understood and constructed, not by means of the verbal descriptions. This should immediately bolster the theorem of convertibility.

(iii) The third field belongs to sign theory because the Kantian claim already includes a tacit controversy. If the sign must have a sensual representation which arises from the sensual understanding of how sensual clues should be related to intellectual meaning or tokens of the mind, then the aggregate thesis becomes weak. Human consciousness must have a powerful tool to secure the relationship between the acoustically and/or visually sensual item and the meaning. This would be rather oscillating or less conclusive if (as the outcome of mere and necessary aggregation) it would have to evolve from simple usage, common convention and settling with custom. Naturally, if human consciousness is capable of determining the mind by way of the senses - a sensual input - then it may rely upon this force in order to fasten the relationship between the sensual item and its role as a sign bearer or regular element of a sign system.

(iv) Finally, the orientation issue (or the fact that human consciousness must rely upon at least two if not three grounds - the real one, its own as the
necessarily subjective one and the objective one - spanning across and implicating every consciousness) is closely related to the projective and perspectivity claim. As is well known, human consciousness must sense and measure space and spatial relationships by way of projection (the subjective red cube included). Accordingly, the perception of shape depends upon the laws of projection or perspectivity, respectively. ‘A thing in itself’ or empirical shapes which translate their impressions must therefore impact upon the human mind in a twofold manner: on the one hand, it involves Euclidian shapes which are known from measurement and geometry. On the other hand, it receives the projective shape in any case, according to perspectivity and the laws of projection. Indeed, both impressions and/or understandings interfere in normal, natural perception of space, where the implied point of view is the clue to intersubjective understanding. In addition - with regard to the normal overlapping (projection) of the past and the future within the present against the continuous flow of unidirectional time (the rational idealization) - this mental situation should be completely complementary or parallel. Now the question arises: why is human consciousness not confused by this fact and how does it manage to cope with this natural (and so far, necessary) alternative? The answer is that the faculties are to separate from their performances and the logical grip on their relation is free to execute the determinative impact. Therefore it is no problem to reserve the rule of spontaneity for the senses, so that they are capable of determining the mind (intellectus). In addition, if human consciousness must perceive according to projection, it must also be capable of being subordinated by the senses: the point of view and the vanishing point are logically of another level or set, because the entire perspectivity is dependent (which is also true of projectivity as the rule of a centered fascicle). So there is another strong piece of evidence to show that the senses (and sensual perception) should not be restricted to mere aggregation but rather the authentic grip of subordination and classificatory implication.

References

Durkheim, Émile 1983. Pragmatisme and sociology. Translated by J.C.Whitehouse. CUP.
______ forthcoming. Metaphysik und die Ordination des Bewusstseins.
Part D:
Value Theory
Paraesthetics:
Irvine School of Aesthetic Theory and Criticism

Ewa Bobrowska, Academy of Fine Arts in Warsaw, Poland

The concept of paraesthetics was developed in the 1980s by David Carroll, the first director of the Critical Theory Institute at the University of California at Irvine. It describes the collaborative research of such distinguished philosophers as Jacques Derrida, Jean-François Lyotard, Fredric Jameson, J. Hillis Miller, Jean-Luc Nancy, and Murray Krieger. It is a new concept of aesthetic theory that favors interdisciplinary strategies and dynamic relations between the aesthetic and the theoretical. This paper discusses the history of its formation and possible meanings of this concept in the context of the postmodern turn towards theory in the United States. The argument concerns the writings of Carroll, Derrida, and Lyotard, in particular: The States of "Theory": History, Art, and Critical Discourse; Paraesthetics: Foucault, Lyotard, Derrida; Discourse, Figure; The Truth in Painting. To reflect upon the democratic, community-based model of the Institute's research, I shall also refer to the contemporary American philosophy of Richard Rorty and his concept of irony.

This paper is a revised version of my book Paratheory.

This paper focuses on David Carroll's concept of paraesthetics. This term should be regarded in the light of the philosophical investigations of the Critical Theory Institute at the University of California at Irvine. The term ‘paraesthetics’ was coined in the 1980s as a result of the collaborative research of the Institute, involving such distinguished philosophers as Jacques Derrida, Jean-François Lyotard, Fredric Jameson, J. Hillis Miller, Jean-Luc Nancy, David Carroll, and Murray Krieger. The Institute as "an interdisciplinary, collaborative research group from various departments in the Humanities and Social Sciences" (cti home) was established in the 1980s, and it is still operating today. The origins of the Institute should be associated with the so-called Yale school of literary criticism - the first centre of deconstruction in the US - and with names such as Paul de Man, Harold Bloom, and Hillis Miller, who later became affiliated with Irvine. In my paper, I would like to point to two terms developed at the time of the Irvine group formation, that is, the notions of paraesthetics and "theory."
In response to the poststructuralist "crisis" of theory, the Irvine research group attempted to shift strictly outlined discipline boundaries in order to analyze the relationship between a work of art and aesthetics, literature and its criticism, or philosophical and linguistic systems. As one of the areas of their investigations, the group identified art analysis in the light of postmodern theory that could reveal new philosophical contexts of artistic activities.

Aesthetics, or rather a specific concept of paraesthetics, constituted an important part of their research. The above-mentioned critics called for the reconsideration of some traditional values, such as aesthetic experience, and for broadening the field of aesthetics so that it could include various disciplines. Moreover, they emphasized a growing interest in anti-art, fringe-art, and the non-aesthetic or extra-esthetic aspects of art. As defined by Carroll in *Paraesthetics: Foucault, Lyotard, Derrida* and in his introduction to *The States of "Theory": History, Art, and Critical Discourse*, the term ‘paraesthetics’ describes a new concept of aesthetic theory that promotes the dynamic relation between the aesthetic and the theoretical. It also undermines the traditional Kantian model of aesthetic autonomy. In other words, it is a strategy of exceeding the boundaries between art and its theory. Thus, paraesthetics begins with the recognition of an aporatic, self-reflexive nature of both discourses: the visual and the philosophical.

Carroll, one of the theorists of the Irvine School, notes that, in response to the latest revisions in art theory, critics should remain open to criticism coming from art itself. Paraesthetics, a term that he proposes, defines a new type of aesthetic theory, for which "art is a question (and in question) rather than a given" (Carroll 1994, 14), as he states. Paraesthetics is an analysis that remains, on the one hand, aware of its limitations, while on the other hand, it draws inspiration and critical models from the territory of art. It is based on an attempt to bridge the boundaries between the language of criticism and the language of art.

In other words, the term ‘paraesthetics’ refers to various post-deconstructive strategies raised by the contemporary critical theory group originally formed at Yale University and later at the Critical Theory Institute, Irvine. The main assumption characteristic of these schools concerns the postulate of exceeding the existing critical modes in order to designate a territory open to an unconstrained interdisciplinary dialogue between the theorists and the practitioners of art. This solution was expected to offer a way out of certain impasses in contemporary aesthetics. Indeed, research collaboration has led to the formation of a new model of theory in the humanities, in its broadest sense, with particular emphasis on the aesthetic, philosophical, literary, semiotic, and sociological strategies of analysis. Their primary objective was to develop a new position which would offer an adequate response to current challenges faced by the humanities and to the need to confront the achievements of individual fields.

However, the difficulties double once the critical discourse embarks on the subject of the postmodern visual arts, due to the ontological distance which
separates art from its theory. As a result, one observes the tendency to discard the representational practices in favor of the poliphony of little narratives in the field of visual arts. The tendency to extend linguistic maneuvers into the visual arts accompanies the dispersion of traditional concepts of beauty to some degree analogous to the postmodern dispersion of meaning. Contemporary audiences are entrapped within self-invading codes that inevitably subvert any straightforward or sincere comment upon the world or themselves. Consequently, the prevailing tendency of the 20th century avant-garde to negate any presuppositions in its persistent quest for ever newer forms of expression, often self-destructive, antiaesthetic, and irrational, calls for adequate categories of analysis.

As a result, critics and philosophers may reach for literary or artistic forms, as was the case with Lyotard, the author of Pacific Wall, a novel which takes place in California and is an extensive commentary on the poetic works of Sam Francis, a painter from Santa Monica, whom the French philosopher met during his stay at the University of California.

In Discourse, Figure, Lyotard claims that painting poses a certain potential for formulating critical statements; at the same time, in the theoretical discourse, there are some traces of the visual. Therefore, paraesthetics should take the place of a frame parallel to a work of art, an inner parergon, to refer to Derrida's terminology. This modern strategy is based on a "constant play between the aesthetic and the extra-aesthetic" (Carroll 1994, 16).

Furthermore, the term 'paraesthetics', introduced by Carroll, could be substituted by or compared to another similar term also mentioned by Carroll, namely, paratheory, which points to the main focus of research of the Irvine group. Paratheory or "theory" (in quotation marks), as it has been used, are broader terms that embrace paraesthetic issues within their scope. However, these terms are mutually dependent; therefore, in an attempt to define paraesthetics, one needs also to consider the Institute's central concept of "theory" and its focus on the recognition of the contemporary state of theory. Thus, the Institute's most important collection of essays, which could be considered their aesthetic manifesto, is entitled The States of "Theory". It was published in 1990 by Stanford University Press, California, and it is composed of two parts entitled Question of History and The Question of Aesthetics. The collection comprises twelve essays by various authors, such as Derrida, Hunt, Leforta, Jameson, Nancy, Krauss, Iser, Krieger, Miller, and Lyotard, and it is prefaced by Carroll's introduction.

Derrida, in his paper included in this anthology, points to the emergence in the United States in the '70s of a very specific, new American use of the term "theory." Like Jonathan Culler, he locates the development of theory in the context of American departments of literature. This theoretical discourse is - according to Derrida - "a purely North American artifact, which only takes on sense from its place of emergence in certain departments of literature in this country. (...) literature and not simply Humanities, if sometimes the Humanities includes historical disciplines" (Derrida 1994, 71). "I would add that this actually happens neither in other departments of this country nor in the
literature departments of other countries in any statistically noticeable way", he concludes (Derrida, 1994, 71). Moreover, Derrida refers to Jonathan Culler's article entitled Criticism and Institutions: The American University. The following quotation from this paper will throw some light on the formation process of the terms ‘paratheory’ and ‘paraesthetics’:

The major critical development of the past twenty years in America has been the impact of various theoretical perspectives and discourses: linguistics, psychoanalysis, feminism, structuralism, deconstruction. A corollary of this has been the expansion of the domain of literary studies to include many concerns previously remote from it. In most American universities today a course on Freud is more likely to be offered in the English or French Departments than in the Psychology Department; Nietzsche, Sartre, Gadamer, Heidegger, and Derrida are more often discussed by teachers of literature than teachers of philosophy; Saussure is neglected by linguists and appreciated by students and teachers of literature. The writings of authors such as these fall into a miscellaneous genre whose most convenient designation is simply "theory," which today has come to refer to works that succeed in challenging and reorienting thinking in fields other than those to which they ostensibly belong, because their analyzes of language, or mind, or history, or culture offer novel and persuasive accounts of signification. (Culler, 1987 p. 87.)

Let me stress again that the research on "theory" and the related concept of paraesthetics began in California at the time of the publication of Culler's paper in 1987 and should be associated with the Critical Theory Institute at the UCI. These are among the most important concepts developed by this group. The terms "theory" or “paratheory” were first used in The States of "Theory" - an important collection of essays, edited by Carroll, by leading figures cooperating with the Institute. On the other hand, the neologism “paraesthetics” is a key term in Carroll's monograph entitled Paraesthetics: Foucault, Lyotard, Derrida.

Carroll, who is currently lecturing at the University of California, has been involved in the research of the Institute from the beginning of its activities. As a specialist on literature, theory and aesthetics, Professor of Romance Languages at UCI and coordinator of the contemporary critical theory research program, he held the position of director of the Institute. He also participated in the first research project initiated by the group, entitled The Aims of Representation: Subject / Text / History. Notably, the international character of this cooperation has been due to the cooperation of not only American and French but also German (Robert Weimann, Wolfgang Iser) and British (Anthony Giddens) scholars. Notably, Carroll is still an active researcher who continues the tradition of the Institute with his lectures and seminars as a part
of the Critical Theory Emphasis, UCI's program, which is open to all doctoral students.

Interestingly, the current ongoing project of the Institute entitled *Poor Theory* seems to refer to the above-mentioned projects from the second half of the '80s. According to its manifesto, poor theory aims to be situated within the framework of interdependence to the concepts of "weak philosophy" of Gianni Vattimo, Grotowski's "poor theater", “architecture for the poor”, “arte povera” by Germano Celant, and the “poor cinema” of Julio Garcia-Espinosa. This project in many ways performs and radicalizes assumptions of the '80s, having excluded to a greater extent than previously the burden of its academic character. It also reaffirms the attitude of openness to new forms of reflection and shows a turn towards new cross-disciplinary solutions located among science, art, and literature. A short quotation from its manifesto will demonstrate the current application of paraesthetic strategies.

Poor theory does not simply celebrate fragmentation and pluralism; rather, it seeks a complex interdisciplinary engagement across cultures, histories, and practices. It draws inspiration and rigor from all disciplines, but it does not seek to redefine theory as a singular disciplinary endeavor. It may be particularly suitable to mingling familiar sites of theory with sites still incompletely engaged by other forms of theory.

The expected outcome of those paraesthetic strategies is the unpredictable, the unknown in Lyotard's terms. Poor theory project exposes also other links to Lyotard and Derrida's reflection, *inter alia*, by drawing attention to the elements of free play and self-irony. In Rorty's vocabulary it would be phrased as a creative, unrestricted way of coping with the reality making use of the imaginative potential, limited only by the interdiction of cruelty.

Similar arguments can be found in Carroll's introduction to the aforementioned collection *The States of "Theory"*, in which he states: "In its critical form at least, theory may be best described as the hybrid and open field, in which the possibilities of the various disciplines and fields, it crosses through and which cross through it, are pursued and experimented with" (Carroll 1994, 3). Each of these interpretations is both a stimulus for further experiments and a contribution to the creative transformations of critical theory itself, as Carroll argues.

Paraesthetic strategies in Carroll's reading of Derrida are accounted for by the term "theoretical jetty," which allows the critic to approach the depth of a work or a process, although it does not ensure control over it. In a conflictual process, jetties compete for a dominant interpretation. Carroll interprets Derrida's jetty figure as an equivalent of research proposals, statements, systems, ideologies, which are related to the term "theory." Each new or "post" theoretical construction struggles for domination, it aims to subjugate previous forms and occupy a privileged, permanent, and safe position. The significance of the concept of theory is further evidenced by Derrida's comment on a joke
based on alternative readings of the phrase "the states of theory" circulating among the staff of the Institute. They would refer to the United States or even to California as the 'state' of theory (a state that differs from other states in the USA by an excess of theorizing).

To sum up, the problem of aesthetics emerged as a significant issue at the very beginning of the program. The aesthetic turn in Irvine took place at the time of the formal foundation of the Institute in 1987. Simultaneously with the establishment of the Institute, there also arose questions about the purpose of formalization of research, the object of which was a turn against all structures. Therefore, one of the assumptions of the Institute was a paradoxical struggle with the petrifying effects of the institutionalization of theory, which should be located in the context of Paul de Man's resistance to theory.

According to Carroll, in the postmodern times, art - as a paradigm that was radically destabilized in the 20th century - has become a discursive phenomenon or a rhetorical trope. On the other hand, writing on art has become an important (para)artistic strategy. Carroll's monograph *Paraesthetics: Foucault, Lyotard, Derrida*, published at the same time as the collection *The States of Theory*, in 1987, also concerns the issue of paraesthetics. With regard to this book, I would like to emphasize his reflection on Derrida and Lyotard.

According to Carroll, paraesthetics - mentioned in the title of the book - is the analysis of art that comes from the nonaesthetic positions. Moreover, to use J. Hillis Miller's metaphor, paraesthetics is a quasi-parasitic form which partially absorbed and consumed its host and consequently took its form, as happened in the paraliterary or parapoetic philosophy of Derrida and Lyotard. Notably, paraesthetics had an effect on transformations of critical theory and should be associated with the decline of the critical school at Yale. This process is often inscribed in the framework of the end of poststructuralism. Carroll notes that analogically to the rococo in art, the late poststructuralist phase of theory is characterized by a variety of critical experimentation strategies and the diversification of attitudes.

As noted by Carroll, the period of increased interest in theory in the United States should be associated with the assimilation of structuralism and poststructuralism that originated in France.

Interestingly, Carroll argues that in America, these tendencies have proven to be more vital and influential than in Europe, at least in France. Despite the exhaustion of theory and the protests of some academic communities, questioning the point of further research that "has become too philosophical, rhetorical," (Carroll 1990, 2), Carroll confirms the importance of critical theory, stressing that a lack of interdisciplinary approaches may lead to a closure within the rigid fields of specialization. The classic model of division into autonomic disciplines may be attractive in its clarity; however, it is not a guarantee of scientific progress. Furthermore, the postmodern turn towards theory in the United States may be, in his opinion, due to the exhaustion of American empiricist and pragmatist traditions.
In this context, Carroll's paraesthetics means opening the critical potential of art and literature and the search for the aesthetic dimensions of theory, as well as the theoretical, conceptual aspects of art. In other words, it is an attempt to reformulate theory by making use of an extremely valuable cognitive distance offered by art. The parasitic relationship between art and theory, to use Miller's metaphor, points to one aspect of this relationship. Art, in response to its theoretical entanglement, can often resort to defense mechanisms. It defends itself against excessive analysis, and paradoxically, the very process of this resistance becomes for Foucault, Lyotard and Derrida a topic of meta-defensive reflection, as Carroll argues.

Derrida's concept of the *parergon* challenges traditional oppositions, such as inside / outside, meaning / form, art / theory, the content/the effect. Carroll cites a very significant passage from *The Truth in Painting*, devoted to the relationship between art and philosophy, in which Derrida states sharply, "[In] reading of these two discourses (...) I notice the following: they both start out from a figure of the circle. And they stay there" (Derrida, 23). Philosophy, which enclosed and displaced the discourse on art, is at the same time enclosed in a circle by art.

Are the elements of art and philosophy not only contradictory but mutually exclusive?, one could ask, and therefore, is the figure of an artist-philosopher possible? In American art, it was exemplified by Joseph Kosuth and Barnett Newman among others.

If Derrida attempts to define aesthetic experience, he does it in negative terms of absence and trace, because beauty or aesthetic experience is not and cannot be complete. As he argues in the chapter *The Sans of the Pure Cut*: "Negativity is significant, working in the service of sense" (Derrida, 1987, 95). Art, according to Derrida, involves the experience of incompleteness, lack, and absence, which is partly compensated for by the critical discourse of aesthetics and art criticism.

Carroll proposes using the critical potential of art and literature as a counterweight or an alternative to more analytic philosophy, which could dynamically shape and verify its arguments. Although he rejects the notion of a dialogue or a consensus present in the philosophy of Rorty, his proposed strategy refers to a model, based on the always precarious balance of opposing factors of art and philosophy. As one of the arguments, he cites the famous Nietzschean motif of domination of philosophy over art, illustrated most clearly in *The Birth of Tragedy*, in the scene where Plato burns his poems to become a disciple of Socrates. The conflict between the Apollonian and Dionysian elements is presented differently by de Man as Dionysian, Apollonian, and Socratic. The American critic notes that the conflict presented in *The Birth of Tragedy* is a simplification. In fact, a semi loser becomes the winner, who wins by the force of influence and intersubordination of both sides of the conflict, in this case, by the transformation of philosophy into a more aesthetically pleasing form and supplementing art with a conceptual element of analysis, as is the case of Kosuth's art. As a result of the confrontation with philosophy, art incorporates the Socratic strategy of the dialogue. While de
Man recognizes the primacy of the literary over the philosophical, Lyotard, as Carroll points out, on the one hand, recognizes the ontological autonomy of art, but also points out that its role is not so much dominant as critical. He argues that "all art, to fulfill its critical function as art, has a critical and self-critical function. It unmasks all attempts to raise any force or entity above the conflict of forces and orders" (Carroll 1987, 27). In a sense, art, according to him, takes over the critical function of theory; all art "to fulfill its critical function as art, [should - E.B] be art and anti-art, at the same time", as Carroll interprets Lyotard's approach from his Driftworks (Carroll 1987, 27).

Most of the argumentation in the first part of the chapter of *Paraesthetics*, dedicated to Lyotard, concerns the possible socio-political entanglements of art (but these considerations, however, are not the subject of this paper). In the summary of this section, there is an interesting diagnosis of the critical function of art in relation to aesthetics. In this light, paraesthetics is not only aesthetic theory with a clearly marked critical and self-correcting nature, but also art is involved in a creative and critical dialogue with theory. Carroll comments in the following way on the new paraesthetic function of art, as revealed in the writings of Lyotard.

Depending on how it is approached, art has very different critical effects in his various works, which makes it possible to argue that the so-called ontological exteriority of art does not determine a space in which the essence of art can be located, but rather posits a distance in which critical alternatives - not just to the historical-political order, but also to the aesthetic realm itself - can be formulated. (Carroll 1987, 30)

In *Discourse, Figure*, two different orders of language and painting are juxtaposed as equal elements of a broader discourse, Carroll argues. Moreover, there is a clear emphasis on the second of these two elements. As Carroll notes, Lyotard considers linguistic forms to be exhausted, determined through the historical and philosophical tradition of self-restraint, with no future potential. He would like to substitute them with the dynamics of painting, which is independent from the philosophical, traditionally linguistic, or semantic grounds, and is open to the unpredictable. Lyotard's figure disrupts discourses and signification and replaces them with intensities. According to Carroll, "It is the realm of movement, difference, reversal, transgression, and affirmation" (Carroll 1987, 31). In *Discourse, Figure*, he claims that painting poses a certain potential for formulating critical statements; at the same time, in the theoretical discourse, there are some traces of the visual. In another passage, inspired by Freudian psychoanalysis, the American critic emphasizes the transformation of both discourse and art into a formless libidal energy. Libidal paraesthetics is basically - as stated - a form of the impossible, or a limit, which aesthetic theory can only approach and indicate, but never phrase.
Carroll briefly discusses also Lyotard's late essay entitled *Adorno as the Devil*, in which the French thinker refers to another tradition, derived from the negative theology, describing attempts to approach the inexpressible or transcendent. Therefore, libidal economics may be considered another form of negative theology. For Lyotard, the libidal structures and art are the highest form of expression, which border on the "inexpressible," comparable to the mechanisms of negative theology. Theoretical and critical discourse, as Carroll admits, should be constantly transgressed and overcome, as is expressed through *via negativa* in apophatic theology.

Moreover, Carroll draws attention to the role of the faculty of narratives or the faculty of narrative imagination introduced by Lyotard. Carroll interprets it as a proposal to extend Kantian trichotomy. To the three cognitive faculties, namely, the intellect, power of judgment, and reason, Lyotard adds an extra faculty of narrative imagination, which "may be defined as the capacity to respond to any narrative with a counter-narrative, to occupy a different place than that assigned by any master-narrative, to improvise and deviate from the assigned plot" (Carroll 1987, 160).

The decision to grant a special role to imagination, or rather to this faculty of the mind, which is manifested through storytelling, links Lyotard's philosophy to Rorty's thought. Both thinkers emphasize the ethical, cultural, and almost universal role of the narrative. This association is not accidental; it confirms the role of literature and its theory for contemporary philosophy, including the philosophical school at Irvine, and the thought of Derrida.

Notably, Rorty's philosophical approach is in many respects parallel to the theoretical perspective of postmodern French thinkers, such as: J. Derrida ad J.-F. Lyotard.

Considering Derrida the most intelligent reader of Heidegger, Rorty *distinguishes* two main issues: theory and irony, stating: "Derrida continues to think about the problem which came to obsess Heidegger: that of how to combine irony and theorizing." (Rorty, 1989: 122) This dilemma played also a significant role in the Institute's research, therefore, critical "theory" may be regarded as a specific form of Rorty's ironical theory. Both approaches are based on constant putting into question, re-description, 'playing off vocabularies,' disciplines, theories against one another, "in a conflictual process of confrontation with the states of the various disciplines and with the states of "theory" itself." (*The States of "Theory" History, Art, and Critical Discourse*, 1990: 22) This idea of philosophy is based on the private, the literary, the imaginative. Therefore, I propose to draw an analogy between Rorty's or Derrida's irony and the Irvine mode of "theory."

Derrida, according to Rorty, relocates the rigid public discourse of philosophy to the domain of the private and, in this way, overcomes the apparent contradiction between theorizing and ironising. Interestingly, this new mode of philosophy is close to the paradigm of literature or artistic fantasy. Rorty comments Derrida' method in following manner:
Such fantasizing is, in my view, the end product of ironist theorizing. Falling back on private fantasy is the only solution to the self-referential problem which such theorizing encounters, the problem of how to distance one's predecessors without doing exactly what one has repudiated them for doing. (Rorty, 1989, 125)

Irony allows for the creation of a private space of distance to the opponent, which encloses free play and contingency of imagination. Theory should remain open to the interplay between rationality and the absurd, seriousness and irony, philosophy and literature, as this approach has it. Consequently, according to Rorty, Derrida's strategy is based, _alter alia_, on unusual reconfigurations. A similar method is distinctive for the Irvine studies and apparent specifically in the recent aforementioned manifesto _Poor Theory_ based on the surrealist manner of associations.

Not only it is based on admiration for possibilities inherent in pluralism, but also on a search for complex interdisciplinary forms of research engagement drawing on various cultures, histories, practices. A similar strategy of playing with the unpredictable defines Rortian imagination paradigms. The dynamics of differences are, according to Rorty, generated by imagination. It is imaginative thinking that allows for transgressing one's final vocabulary and for strengthening community ties.

This comparison is further reinforced by re-reading Rorty's interest in Derrida's philosophy with an emphasis on the role of literary criticism. Both Derrida and Rorty, as well as Hillis Miller, Paul de Man, and other theorists from the Irvine circle address the concept of 'philosophy as a kind of writing' or a literary genre, based on a corresponding dialogic, Socratic, and ironic mode of philosophical dispute. Moreover, Rorty's cultural criticism which involves literary and aesthetic theory and criticism, philosophical thought, ethics, social theory, etc. may be parallel to Carroll's concept of paraesthetics: a new versatile, comprehensive, and flexible intellectual discipline, rooted in the antique tradition of a dialogue. (Lyotard 203). In Plato's dialogues the interlocutors do not reach any conclusive universal. However, it is the sole act of the dialogue, circulation of thoughts between the truth searching philosophers, that could become a model of any teaching or research process in the humanities, which however can never offer any exclusive solution. The postmodern strategy of research may be well described by the categories drawn from Rorty's writings such as: curiosity, negotiating abilities, antiessentialism. Furthermore, the theory as described by David Carroll is in many respects analogous to Rorty's concept of irony. The critical reflection of the Irvine critics and philosophers fulfills all three criteria of ironism described by Rorty. First, in the manner similar to the Rortian ironist, the Californian theorists experience strong and constant doubts regarding the final vocabulary they use. In the introduction to _The States of "Theory"_ Carroll notes that the members of the Institute should remain alert to the petrifying effects of institutionalization of theory. Critical theory, similarly to Rorty's final vocabulary, should be
Paraesthetics: Irvine School of Aesthetic Theory and Criticism

constantly revitalised, created anew, to play off old with the new. Criticism just as Rorty's irony is to become a guarantee of the eternal youth of any project. Conflicts, diversity allows for still new re-descriptions, situating books and protagonists in new circumstances. Critical theory - according to Carroll, just as irony is based upon constant uncertainty and doubts, formulating new questions that have not been yet formulated, in the process of self-questioning typical of the Rortian ironist. Like the Californian ironists, the ironist is fascinated by conflicts generated by heterogeneous elements. In other words, the Rortian ironist would feel at home at the UCI, which in some respect has become a place of implementation of Rorty's irony. In this light, critical theory may be regarded as a form of new ironical theory advocated by Rorty.

Theory as irony wants to remain destructively creative, wants to struggle at the borders and frame edges. Both ironists and paraesthetians read literary critics, because they have a large circle of acquaintances, so they can become better moral advisers.

Both ironists and paratheorists share the predilection for deconstruction, as well as the method of research based on description, comparison, dialectics, and most of all irony, language games, rhetorics, contrast and apories. However, the criteria of novelty and creativity remain the most important. Therefore Rorty's term modern intellectuals seems adequate also to describe the paratheorists. Last but not least, common to both concepts is the fusion of ethic and aesthetic traits, especially apparent in Derrida and Lyotard.

Such reflection aims not as much at conclusiveness as at originality. Why to write this way? Asks Rorty. May philosophy which discloses its fascination with literature, rhetorics, and imagination still be classified as philosophy? Rorty's and David Carroll's answers are similar: theory or research is worth pursuing as long as it attempts to transcendent rigid self-imposed criteria, as long as it remains critical or ironic, that is creative with regard to its own form. The criteria proposed by Rorty: of novelty and originality further reinforce the mutual entanglement of philosophy and literature. Not surprisingly, Rorty uses the term reading philosophy and compares Derrida's philosophy to Marcel Proust novels.

The third criterium proposed by Rorty involves considering literature a medium for compassion and forming ethic attitudes. The commitment on the side of suffering locates Rorty's thought close to Lyotard. Common to both philosophers is the coexistence of three important elements: philosophical, aesthetic, and ethic.

Furthermore, the predilection for plurality and the respect for irreducible differences or aporias constitute the common denominator of Derrida's concept of différence, Lyotard's notion of différend, and Rortian understanding of community (also research community). Listening attentively to variety of languages is an acknowledged method of research because "the being speaks in a variety of ways, everywhere and continuously through every language" as Derrida concludes.

The unique formula of intellectualism played a significant role in the Institute's strategy. Notably, it was parallel to Rorty's understanding of ironic
An Anthology of Philosophical Studies
Volume 8

intellectualism. In both cases the metaphysical tradition based on seeking for the truth, was replaced by, more liable to the ironic method, discussion about philosophy, art, and literature enriched with some ethical undertones. The aim of the critic would not be hermeneutics, but rather "placing books in the context of other books," comparing and playing off texts or their protagonists. This method ascribes literature, theory and philosophy in the universal dynamics of variability and recurrence, in other words in the dynamics of the Derridian différence. The potential of knowledge and intelligence involves the art of playing with the tradition, culture, literature by relocating, reinterpreting, differing, as Poor Theory manifesto makes it explicit.

The analogous mode of philosophical discussion refers to ironic dialectics based on the act of reinterpretation. As the best practitioners of this strategy Rorty mentions: Hegel, Nietzsche, Heidegger, Derrida. Respectively, one may search for the origins of the Institute's research strategy based on confronting diverse arguments in Lyotard's thought and in dispersion based upon Kant's division of the faculties.

The search for one prevailing structure or a universal dictum which could encompass all possible varieties of aesthetic and antiaesthetic realms encountered in contemporary art is doomed to failure, states Lyotard. In Lyotard and Carroll's rhetoric, conceding defeat marks the shift toward a new opening that validates certain strategies of crossing the established interdisciplinary boundaries. Furthermore, it is within the realm of paraesthetics that one can encounter similar endeavors to relocate the restrictive disciplinary demarcation lines.

Thus, the research undertaken within the Critical Theory Institute may be defined as a process of collaborative, maieutic para-theorizing, the process of questioning which is never satisfied with the received response, regardless of the (current or Californian) state of theory.

References


Abstract In this paper, I consider three standard interpretations of the standard of the reasonable person: the customary view, the avoidability view, and the indifference view. In a comprehensive treatise, Mayo Moran contends that the indifference view is required to eliminate prejudiced interpretations of SRP, which have plagued judicial decision-making since O.W. Holmes originated this standard in 1881. Here, I dispute this view, and argue that the avoidability view is supported by a range of familiar moral principles. I also contend that we can mitigate misapplications of SRP by introducing and enforcing procedural constraints, as opposed to the substantive constraint on decision-making that Moran urges.

Key words Reasonable person; responsibility; liability; avoidability; Kant; utilitarianism

Introduction

O.W. Holmes’ standard of the reasonable person (SRP) has been criticized by feminists, critical race theorists, and critical disability scholars for defining ‘reasonableness’ by reference to the traditional ideal of the White, middleclass, able-bodied male, thereby marginalizing and unfairly penalizing anyone outside of this privileged group. In this paper, I consider three interpretations of SRP—the customary view, the avoidability view, and the indifference view—and argue that the avoidability view is corroborated by insights in moral philosophy and a combination familiar moral principles. I then argue that in light of this interpretation, the avoidability view manages to avoid the charge of discrimination, and actually fares better in this regard than its main rival, the indifference view.
Three Interpretations of SRP

SRP plays a central role in determining culpable negligence in the common law. It was originally advanced by O.W. Holmes in his seminal ‘The Common Law’ (1881) primarily as a means of dealing with cases of negligence and recklessness, but with broader applications as well. SRP dictates that responsibility is a matter of ‘what would be blameworthy in the average man, the man of ordinary intelligence and reasonable prudence’ (p. 46). The justificatory basis for this standard is twofold: First, due to ‘the impossibility of nicely measuring a man’s powers and limitations,’ subjective tests of liability (which focus on the agent’s subjective mental states) are unfeasible; and secondly, ‘when men live in society, a certain average of conduct, a sacrifice of individual peculiarities going beyond a certain point, is necessary for the general welfare’ (p. 46). The first principle is articulated in terms of fairness or desert, and the second in terms of social utility. In this way, SRP preserves a just balance of individual rights and liberties against ‘the correlative amount of [social] security people can reasonably expect’ (Holmes, p. 174).

In spite of the initial plausibility of this view, SRP has unfortunately been applied in ways that discriminate against vulnerable political groups. Mayo Moran, who has devoted a book to this issue (2003), argues that SRP’s dependence upon prevailing notions of ‘normalcy’ and ‘naturalness’ has served to rationalize ‘the unequal treatment of both the developmentally disabled and women’ (p. 147). For example, SRP has been used to oppress persons with intellectual disabilities, to exclude women from professional and public life, and to impose harsher sentences on women than men (Moran, pp. 147-148). Moran proposes to resolve this problem by redefining ‘the reasonable person’ so that it includes certain biographical and descriptive qualities of the defendant (such as age, intelligence, and physical ability), while excluding specifically ‘prudential or normative shortcomings’ (p. 243). This definition corresponds with Anthony’s Duff’s earlier ‘indifference account’ (1990), which holds people culpable for ‘only those actions that betray indifference to the interests of others’ (Moran, p. 12). One of the distinctive features of this view is that it does not excuse defendants on grounds of moral ignorance or insanity.

Moran compares this interpretation against two common alternatives. The first is the ‘customary view,’ which defines negligence in light of community standards. Specifically, it holds a person liable for negligence if she fails to conform to our expectations of what ordinary members of the community would do in similar circumstances. Moran rejects this approach because, ‘though it may coincidentally embody respect for equal moral standing, it will not necessarily do so’ (p. 247). That is, in an inegalitarian society it may simply reinforce customary discrimination, which would undermine the central egalitarian function of the justice system. Moran believes that we need a theory that isolates and condemns prudential and normative failings per se.
The second theory is the ‘avoidability view’ popularized by H.L.A. Hart (1970), which defines legal liability in terms of whether the defendant had, ‘when [she] acted, the normal capacities, physical and mental, for doing what the law requires and abstaining from what it forbids, and a fair opportunity to exercise those capacities’ (Hart, p. 152, quoted in Moran, p. 243). Moran rejects this account on essentially the same grounds as the customary view, i.e. because it does not rule out specifically prudential and normative deficits.

Moran goes on to defend the indifference account because she believes that it is the only theory capable of protecting the interests of political minorities by ruling out discriminatory interpretations of reasonable personhood. However, I believe that she may have overlooked the advantages of this view by focusing too narrowly on its (contingent) historical misapplications, and ignoring its ability to explain a central and indispensable component of culpability. In the next section, I argue that if we can restrict misconceptions of SRP by imposing procedural constraints on judicial discretion, the avoidability view represents a highly eligible means of determining responsibility and legal liability.

The Customary Account, the Avoidability Account, and the Indifference Account

In considering the customary account and the avoidability account, the first thing to notice is that the two are not easily separated into distinct categories, inasmuch as customary behaviour naturally and inevitably informs our conception of what it would be reasonable to expect of a person under duress. Specifically, it informs our conception of how much it would be fair to expect ‘a person of reasonable firmness… to resist’ or avoid, as required in The Model Penal Code for a duress excuse (p. 367). In John Doris’ words,

such observations [as that torture is excusing but embarrassment is not] should not be taken to indicate that excuses are determined simply by population ‘base rates’; observing ‘everyone’s doing it’ does not have the makings of an excuse. Still, reflection on base rates helps determine what can be expected of a particular individual in particular circumstances; surely it is partly because most people yield under torture that it seems unfair to hold victims liable for failing to resist it (2007, p. 527).

Hence, customary behaviour is notionally linked with avoidability insofar as it naturally informs our understanding of when an action was preempted by duress. Nonetheless, there must be allowances for saying that certain practices or patterns of behaviour are blameworthy even though they are widespread, or else the law cannot fulfill what H.M. Hart describes as ‘one of the prime functions of the manifold safeguards of American criminal procedure,’ i.e.
'enhancing a sense of security throughout society' (1958, p. 401). One natural way of achieving this goal is to excuse people only if their action was demonstrably *unavoidable*, in some reasonable sense of the word (to be considered shortly). Thus, the avoidability view can be seen as an improvement on the customary view, in that it respects the excuse of duress while simultaneously reinforcing one of the main functions of criminal law. It holds people accountable for customary wrongdoing when the person could not reasonably have done otherwise. In this way, it is midway between the customary view and the indifference view, which imposes a stricter standard of liability.

Now, the main difference between the avoidability view and the indifference view, as we saw earlier, is that the former excuses people for actions stemming from inexorable moral deficits, whereas the latter does not. This is because the indifference view holds people accountable for indifference no matter what the underlying deficit or degree of disability, whereas the avoidability view excuses indifference when it is due to uncontrollable deficits which significantly impair volitional control or behavioural regulation. In other words, it treats uncontrollable moral deficits as a type of duress. What reason do we have to accept this view? If we turn to moral philosophy, we do not need to look far for avoidability-type accounts of responsibility. These accounts may offer *prima facie* evidence for the same type of excuse in the common law, inasmuch as legal accounts ostensibly invoke and hinge on moral justifications. (More on this at the end of the section). I propose that we consider Susan Wolf, Miranda Fricker, Gideon Rosen, and Michael Slote as examples of avoidability-based views that have relevance for legal thought.

Wolf (1987) defends a well-known avoidability-type theory called the ‘sane deep-self view,’ which holds that in order to be morally responsible, an agent must satisfy three criteria: (a) her actions must be within the control of her will, (b) her will must be within the control of her deep self, and (c) her deep self must be sane. The third condition entails that the agent must have the capacity ‘cognitively and normatively to understand and appreciate the world for what it is’ (p. 387). This view constitutes an avoidability account because it defines moral insanity in terms of an agent’s ‘lack of control at the deepest level’ (p. 384), which is categorically different from ordinary moral confusion and mistakenness, and as ‘unavoidable’ incapacitation in one’s ability to grasp moral reality (p. 384). This differs from the ‘plain deep-self view,’ which requires only conditions (a) and (b). The latter is essentially a form of indifference account, since it holds agents responsible for indifference regardless of the cause and degree of moral deficit.

Wolf’s theory hinges largely on the intuitive appeal of a well-known thought experiment. Specifically, Wolf asks us to envision JoJo, the favorite son of Jo the First, an evil and sadistic dictator of a small, undeveloped country. Because of his father’s special feelings for the boy, Jojo is given a special education and is allowed to
accompany his father and observe his daily routine. In light of this treatment, it is not surprising that little JoJo takes his father as a role model and develops values very much like Dad’s. As an adult, he does many of the same sorts of things his father did, including sending people to prison or to death or to torture chambers on the basis of whim. (p. 379)

We are to assume that JoJo wholeheartedly endorses his sadistic actions upon critical reflection. ‘In light of JoJo’s inheritance and upbringing—both of which he was powerless to control,’ says Wolf, ‘it is dubious at best that he should be regarded as responsible for what he does’ (p. 380). Insofar as JoJo’s formative circumstances were ‘out of his control,’ and in this sense unavoidable, we cannot regard him as morally responsible.

In a similar vein, the virtue ethicist and feminist epistemologist Miranda Fricker defends an avoidability type of account which excuses people for culturally-induced moral ignorance (2007). In particular, Fricker argues that a person ‘cannot be blamed for making a routine moral judgment. But one can none the less be held responsible for making a merely routine judgment in a context in which a more exceptional alternative is, as a matter of historical possibility, just around the corner’ (p. 105). She gives the example of Herbert Greenleaf from The Talented Mr. Ripley, who discounts the testimony of his son’s girlfriend on sexist grounds. Fricker says that Greenleaf should be excused for his prejudiced thoughts and actions because he was not ‘in a position to know better’ (p. 100). That is, as a 1950’s-era Midwestern American man, he did not have access to a more enlightened perspective. However, since moral alternatives are clearly available to most present-day Westerners, the same excuse does not apply to us (barring very abnormal circumstances). Hence, the excuse of moral ignorance is limited to peculiar cases.

Along the same lines, Michael Slote (1982) contends that ancient slaveholders should be excused for their transgressions inasmuch as they were ‘unable to see what virtue required in regard to slavery…, not due to personal limitations (alone) but [also by virtue of] social and historical forces, by cultural limitations’ (p. 72). And Gideon Rosen (2003) argues that the Hittites should be excused for practicing slavery because, ‘given the intellectual and cultural resources available to a second millennium Hittite lord, it would have taken a moral genius to see through to the wrongness of chattel slavery’ (p. 66). This reflects a predilection in moral philosophy to recognize avoidability as a factor in moral responsibility, based on the agent’s specific epistemological location. Arguably, these accounts rely on the same kinds of considerations that we find in Holmes’ ‘The Common Law,’ even if they do not explicitly invoke the reasonable person. Rather, these philosophers tend to rely on the evidence of ‘our pretheoretical intuitions’ (Wolf, p. 382), ‘widespread intersubjective agreement’ (Wolf, p. 386), persuasive thought experiments and anecdotes.
While these moral justifications have elicited much sympathy, one might object that they rely too heavily on anecdotal evidence, emotional appeals, and *ad populum* reasoning. It is reasonable for philosophers to demand a justification for the excuse of avoidability that does not hinge on intersubjective agreement. This has been a trenchant objection in contemporary analytic philosophy, and one standard way of resolving it, at least since Rawls (1951), is by using the method of *reflective equilibrium*, whereby we integrate our considered judgments or intuitions about particular cases, theoretical insights, and moral principles into a coherent whole. (For an extensive description and defense of this general method, see Daniels 1979). While we have seen intuitive and theoretical arguments which speak in favour of the avoidability view, there are also moral principles embedded in Holmes’ defense of SRP that bear consideration. Holmes, recall, says that the law should measure responsibility based on ‘what would be blameworthy in the average [person], the [person] of ordinary intelligence and reasonable prudence’ (p. 51). (As per common practice, I have substituted gender-neutral language for traditional pronominal usage. The central meaning of the text and moral justifications remain the same). While promoting ‘the general welfare,’ Holmes simultaneously grants clemency to those suffering from inordinate or unusually severe deficits, such as immaturity and ‘madness’:

[Legal standards] do not merely require that every man should get as near as he can to the best conduct possible for him. They require him at his own peril to come up to a certain height. They take no account of incapacities, *unless the weakness is so marked as to fall into well-known exceptions, such as infancy or madness*. (p. 50, emphasis mine)

The natural interpretation of these competing claims is to see them as two sets of familiar principles: consequentialist and deontic ones. On the one hand, requiring people to come up to the standard of the majority is likely to compel them to do their best. This accords with the familiar pragmatic/prospective function of responsibility ascriptions, i.e. to encourage people to foster as much normative competence as possible for the common good. On the other hand, Holmes’ mitigating attitude toward those with severe deficits reinforces the backward-looking/retrospective function of responsibility ascription, i.e. to treat individuals fairly. This latter concern is particularly explicit in the following passage:

There are exceptions to the principle that every man is presumed to possess ordinary capacity to avoid harm to his neighbors, which illustrate the rule, and also the moral basis of liability in general. When a man has a distinct defect of such a nature that all can recognize it as making certain precautions impossible, he will not be held answerable for not taking them. A blind man is not required to
see at his peril; and although he is, no doubt, bound to consider his infirmity in regulating his actions, yet if he properly finds himself in a certain situation, the neglect of precautions requiring eyesight would not prevent his recovering for an injury to himself, and, it may be presumed, would not make him liable for injuring another. (p. 109)

Although Holmes resists wholesale appropriation of Kantian ethics, this passage shows that he nonetheless endorses a form of deontic reasoning, which is meant to be tempered by consequentialism. This is clear in his (qualified) objection to Kant, where he contends that although Roman law is ‘profoundly influenced if not controlled by some form of Kantian or post-Kantian philosophy’ (p. 206), contemporary English law diverges from Kant in recognizing that individual rights sometimes yield to the greater good. Thus, he holds that ‘no society has ever admitted that it could not sacrifice individual welfare to its own existence’ (utilitarian principle); and yet ‘no civilized government sacrifices the citizen more than it can help, but still sacrificing his will and his welfare to that of the rest’ (deontic constraint) (p. 43). This reflects Holmes’ commitment to a just balance of deontic respect for individual rights/autonomy against the state’s right to preserve social utility at the farthest margins of political need. When social disintegration and political instability are at stake, social utility becomes a guiding principle. While there are different ways of balancing these values, the important thing is that both must be acknowledged, to a non-negligible extent, in a just social constitution. This perspective is also reflected in contractarian theories such as Rawls’ ‘Theory of Justice’ (1971), which presents two principles of justice: (i) the equal liberty principle, which states that each person must ‘have an equal right to the most extensive basic liberty compatible with a similar liberty for other’ (p. 615), and (ii) the difference principle, which states that social and economic inequalities must ‘be arranged so that they are both (a) reasonably expected to be to everyone’s advantage, and (b) attached to positions and offices open to all’ (p. 53). The second constraint prevents the equal liberty principle from enforcing a destructive, hedonistic utilitarianism that would ignore ‘the wellbeing of society’ (Rawls, p. 53).

Now, while the avoidability view, as articulated here, has the support of epistemologically-minded moral philosophers and Rawls’ two-principle approach, it is important to construe this notion in the right sense, so that it achieves the right balance of fairness and social welfare. Since Kant has advanced a particularly influential account of avoidability in his discourse on ought-implies-can—that is, the notion that in order to be held responsible for an action, the action ‘must be possible under natural conditions’ (GW, p. 540)—this is an ideal starting place for the present inquiry. Unfortunately, the qualification of ‘natural conditions’ is eminently vague and contestable, rendering the correct interpretation of ‘possibility’ equally elusive. How should we understand this modal operator? Surely the logical sense of possibility is
too strict, as it holds one responsible for obeying the law unless doing so entails a logical contradiction; that is, the very idea of obeying the law must be impossible. This would fail to excuse anyone. Physical possibility is also too stringent, since although it may excuse people who break the law due to physical disability, it fails to account for cases of necessity, duress, coercion, and the like, where strict legal compliance is psychologically untenable or prohibitively difficult. To adapt an instructive example from Doris (2007), if someone is coerced into lying under oath by an intimidating mafia agent who threatens to break both his legs, the person is entitled to an excuse of duress, even though disobedience was physically possible. (That is, he could have acquiesced to a severe beating). We do not hold people to this strict a standard of legal compliance because we recognize that it would be unfair and self-defeating—unfair because it imposes an unrealistic burden on people, and self-defeating because it is useless against the threat of serious and immediate bodily harm implied in the definition of duress. In light of this, we need a different conception of possibility, which fits with the Model Penal Code’s definition of duress as submission to what ‘a person of reasonable firmness in [the defendant’s] situation would have been unable to resist’ (Doris 2007, p. 526), and which accommodates human limitations.

I submit that a suitable candidate can be found in Fricker’s account of responsibility, which offers an epistemic reading of ‘possibility.’ Fricker holds that

one cannot be blamed for doing something if one was not in a position to access the reason to do it. This presents a special case of “ought” implies “can,” since in our example the “can” part is a matter of whether Greenleaf could reasonably be expected to achieve the critical perspective on gender that would have enabled him to question his [implicit bias] (p. 101).

Since Greenleaf (ex hypothesi) could not, he is to that extent not responsible. (There may be degrees of responsibility, which will naturally correspond degrees of epistemic constraint. Evaluating the degree of exculpation will be a matter of discretion). On the other hand, Fricker emphasizes that responsibility applies if ‘a more exceptional alternative is, as a matter of historical possibility, just around the corner’ (p. 105). This interpretation fits nicely with Holmes’ dual view of responsibility, which integrates deontic concerns about fairness with utilitarian solicitude for the common good. Far from endorsing moral apathy and cynicism, Fricker is eminently concerned with fostering epistemic responsibility, which can be achieved by tailoring responsibility ascriptions to a person’s capabilities. This moral-capacity view undergirds her positive virtue ethical program, which offers advice for fostering epistemic virtue, such as valuing accuracy, sincerity, and fair-mindedness (pp. 111-18). One might object that this standard is too permissive or lackadaisical, but it finds support in empirically-informed moral
psychology, such as Owen Flanagan’s “Varieties of Moral Personality: Ethics and Psychological Realism” (1991). There, Flanagan defends the Principle of Minimal Psychological Realism (PMPR), which holds that a ‘regulative morality’ must ‘draw on an image of ourselves that we are capable of admiring and to which we can in some sense imagine conforming. A normative conception which fails to meet certain standards of psychological realizability will fail to grip us, and in failing to grip us will fail to gain our attention, respect, and effort’ (1991, p. 26). This corroborates an epistemological reading of ought-implies-can, and thus of ‘avoidability.’ Responsibility, if it is to comply with basic deontic and pragmatic requirements, must be targeted to agents who are minimally morally competent.

This interpretation also fits with neoKantian readings of ought-implies-can. For instance, Dana Kay Nelkin (2011) states that Kantian responsibility entails action-directedness, inasmuch as ‘action-directedness is built into the very idea of obligation’ (p. 114). That is, moral responsibility ascriptions presume that the target agent possesses the ‘perceptual, cognitive, and emotional’ capacities required to respond to ‘ought’ claims (Nelkin, p. 76). This view logically rules our irredeemably morally insane and disabled agents. In a deeper analysis of Kantian scholarship, Darwall notes an inconsistency between the Groundwork, which assumes that responsibility is first-personal, and the Critique of Practical Reason, which endorses a second-personal view, according to which an action is obligatory only if the agent possesses an internal reason, or ‘source of motivation to do as he was morally obligated’ (p. 34). Darwall endorses Kant’s ultimate decision ‘to abandon the Groundwork’s strategy’ (p. 32) and adopt an internal reasons view. Thus, while an alternative interpretation of Kant is available, it is inconsistent with Kant’s final verdict, as well as the need for a minimally motivating ethical theory. This interpretation of Kant fits with Holmes’ explicit concern for the rights of persons with self-regulating ‘infirmities,’ ‘madness,’ and ‘infancy,’ whom he excuses from responsibility. It also supports an avoidability interpretation of SRP which excuses such agents.

With all this in mind, if we take a reflective equilibrium approach to justification, we find that these arguments drawn from disparate sources collectively justify an epistemically-informed avoidability interpretation of SRP. We have seen that avoidability figures prominently in several philosophical accounts of moral responsibility that rely on pretheoretical intuitions and intersubjective agreement, in classical constructarian theories which support a deontic constraint on bare hedonistic utilitarianism, and in neoKantian scholarship which supports an action-directing conception of ought-implies-can. These accounts hang together in a mutually-reinforcing theoretical framework for the avoidability view, with avoidability defined in terms of capacity-limiting epistemic constraints.

There still remains the question alluded to earlier of whether moral arguments are relevant to a legal understanding of SRP. There is reason to think that they are. In the first place, many legal philosophers believe that the
criminal law aims to classify kinds of moral wrong and to punish those who commit them, suggesting that there is a conceptual overlap between the two fields (e.g. Moore 1997, Tadros 2011). But even if one does not subscribe to this strong view, it is fairly uncontroversial that ‘thin’ moral concepts—such as justice, fairness, and social utility—transcend domains. Holmes’ account makes open and unapologetic use of morally-laden terms such as fairness, the general welfare, rights, and obligations, and this is typical of legal scholarship. Holmes states, for instance, that ‘the law must give a man a fair chance to avoid doing the harm before he is held responsible for it’ (p. 144); that the state can restrict individual liberty if necessary ‘to increase the general welfare’ (48); and that a person has a ‘right’ to form a contract with another person, which then creates an ‘obligation’ to respect the resulting agreement (p. 333). This indicates that these rationales are bilaterally valid. Their force, however, may differ according to the particularities of the case. For instance, deontic considerations may weigh more heavily in legal deliberations if the prospective legal sanctions are particularly severe; but this should not be seen as a categorical difference, since moral responses can also impose heavy burdens. Thus, while the same considerations apply across contexts, their weightiness may vary according to circumstances. The important lesson for our purposes, though, is that moral considerations are applicable to legal liability, and thus to evaluating SRP. While the force of utilitarian versus deontic concerns may be issue, their existence as grounding principles for SRP should no longer be in question.

Response to the Charge of Discrimination

Moran’s main concern is that the avoidability view discriminates against political minorities. My response is that while SRP has been misapplied in the past, this is not due to avoidability in and of itself; rather, it is due to a misapplication of SRP based on the judge’s or prosecutor’s subjective biases and prejudices. As a theoretical approach, the avoidability view is no more vulnerable to abuse than competing theories, including the indifference view, and it has the distinct advantage of treating people fairly in light of their capacities.

Arguably, SRP has been misapplied mainly because prosecutors and judges misunderstood what it would be reasonable to expect of different types of defendants. Note that this does not necessarily reflect a misapplication of ‘avoidability,’ since it is not clear that the judges’ decisions were guided by this principle as opposed to another. They could have been (mis)employing a customary or indifference view, or using some other deliberative consideration. Moreover, even if they were deciding on the basis of avoidability, they may have misunderstood the meaning of this concept, which is a problem that can in principle be rectified using procedural methods. Since judicial discretion is important for the purposes of considering and weighing the unique features of
each case, we should be wary of substantive methods that rule out certain principles categorically. For these reasons, I favour the use of procedural constraints as opposed to substantive ones for ensuring a correct and non-discriminatory application of SRP. Although I cannot exhaustively describe these conditions in the short space remaining, they include the following: appointing more political minorities to the bench; teaching law students the principles of feminist jurisprudence, critical race theory, and critical disability theory; incorporate more expert testimony and interdisciplinary evidence into the courtroom; following Keith Culver’s (2008) advice, of attending ‘to history and social science for evidence that law, legal concepts, and legal institutions contain biases against certain groups or interests’ (p. 212), and noting ‘the substantive effect of law more than its formal appearance’ (p. 213); and generating more responsible scholarship on the subject. Unlike Moran’s proposal, these safeguards do not exclude the role of avoidability as one of the possible considerations involved in determining responsibility, but they place constraints on all forms of judicial decision-making by introducing education and diversity. These measures will facilitate a correct interpretation of SRP across all contexts.

Finally, if my interpretation of SRP is correct, then Moran’s view actually enforces a distinctive form of prejudicial discrimination: namely, discrimination against individuals who are incapable of responding to moral reasons. According to the previous analysis, it is both illogical and unfair to hold a morally insane person responsible, just as it would be illogical and unfair to hold a blind person responsible for causing harm due to visual disability. Moran’s view rejects this principle, and in so doing, contradicts several established legal defenses, particularly the mental disorders defense (or the insanity defense in the American law), and the cultural defense. In Canadian and American law, ‘it may be possible in some cases to introduce cultural factors into court under the rubric of the insanity defense’ (HLRA, p. 1294). In particular, the defendant may argue ‘that his cultural values were so different from the majoritarian values reflected in the criminal law that “he lack[ed] substantial capacity either to appreciate the criminality [wrongfulness] of his conduct or to conform his conduct to the requirements of law”’ (HLRA, pp. 1294). This reasoning hinges on the viability of avoidability as an excuse. Moran’s view, in effect, would banish these two entrenched legal defenses, yet she nowhere explicitly defends this substantial change to the law. If we are to take her view seriously, we need a defense for this radical revision. Yet even if one is offered, there are still strong reasons for permitting avoidability to play a role in deliberations about responsibility and legal liability.

References


Plea Bargaining’s Moral Controversies

David Heise, Humboldt State University, USA

In this paper, I examine four challenges to the moral status of plea-bargaining:
1) Is administrative efficiency the primary motivation/justification for plea bargaining? 2) Is plea bargaining analogous to a contract made under duress and therefore unfair? 3) If a defendant facing a felony charge elects to stand trial instead of pleading guilty in the expectation of a more lenient sentence, upon conviction, is he/she being punished more severely for having exercised his/her constitutional right to trial by jury? 4) Does the process of plea bargaining dispense systemic (as opposed to aberrational) injustice? My conclusion is that ultimately plea bargaining, while in need of reform, is indeed a moral practice within the contemporary judicial system.

Keywords: Plea-bargaining; ethics; coercion.

Introduction

Plea bargaining, once the exception to trial by jury, has come to dominate the administration of justice in the United States (and is on the rise in many other countries). The verdict by judge or jury has been predominantly replaced by the negotiated plea. The Sixth Amendment to the United States Constitution guarantees every citizen the right to be judged by a jury of his or her peers. Yet most people who face criminal charges choose to give up that right and plead guilty. Under the process of plea bargaining, defendants agree to waive their constitutional right to trial by jury and plead guilty in exchange for a lesser sentence or a reduced charge. Plea bargaining involves negotiations between defense counsel, on the part of the defendant, and the prosecution, on the part of the state, regarding the conditions under which the defendant will enter a plea of guilty. Today, the vast majority of criminal cases are settled in this fashion.

The ascendancy of plea bargaining has prompted serious moral questions and criticisms regarding the process, leading some to call for reforms to the process and others to call for its outright abolition. The former maintain that there is nothing wrong with the process of plea bargaining itself, but that it is susceptible to abuses (as is any process) which can be avoided via proper guidelines and provisions. The later maintain that plea bargaining is inherently
unjust and thus cannot be reformed. As Kenneth Kipnis has rightly noticed, ‘though philosophers do not often treat issues arising in the area of criminal procedure, there are problems here that cry out for our attention.’ (Kipnis 1992, p. 245) In order to examine the moral status of plea bargaining, this paper focuses on four questions: 1) Is administrative efficiency the primary motivation/justification for plea bargaining? 2) Is plea bargaining analogous to a contract made under duress and therefore unfair? 3) If a defendant facing a felony charge elects to stand trial instead of pleading guilty in the expectation of a more lenient sentence, upon conviction, is he/she being punished more severely for having exercised his/her constitutional right to trial by jury? And, 4) Does the process of plea bargaining dispense systemic (as opposed to aberrational) injustice? My conclusion is that ultimately plea bargaining, while in need of reform, is indeed a moral practice within the contemporary judicial system.

Is Administrative Efficiency the Primary Motivation/Justification for Plea Bargaining?

Plea bargaining has arisen in an era where a flood of criminal cases are coming before the limited resources of the court system. The state is overwhelmed with cases and does not have the resources to bring them all to trial. Often, in borderline cases, there is not sufficient evidence to be confident of a jury’s conviction. With taxpayers unwilling to supply the enormous increase in funds required to provide the courts and personnel needed to bring all cases to trial, the law of supply and demand has taken effect: ceteris paribus, if there are too many cases for the court system to handle, then many will have to be settled relatively quickly and inexpensively outside of trial, via negotiated pleas.

While it is true that administrative efficiency, or response to a growing backlog of cases, is both the reason plea bargaining came into widespread use and one of the primary reasons it still occurs, it should be understood that administrative efficiency is neither the cause of plea bargaining’s origination, nor the sole motivation/justification for its current widespread use. Plea bargaining existed as a seldom used legal option long before the modern court system was forced to engage an over burdensome docket. Once plea bargaining came into widespread use prosecutors and defense council realized that they had a powerful new means to meet their respective ends, creating alternative motivations and justifications for the current use of plea bargaining. Consider, for example, one such justification described by Judge Thomas W. Church Junior:

Whether a particular defendant will be found guilty at trial, however, is subject to the uncertainties of judicial rulings on admissibility of evidence, the ability of opposing council, and the unpredictability of
the jury. A conscientious prosecutor, mindful of his responsibility to protect the public welfare, might rationally conclude that the certainly of a lower sentence might better serve the public than the risk of acquittal at trial (Church 1992, p. 263).

The logic at work here is that in cases where the prosecution’s case is not the best it could be, and the probability for conviction is not high, it would be better to convict defendants of a lesser charge via a plea bargain than to chance total acquittal at trial. Likewise, defense council may invoke this logic in the reverse direction and reason that it is better to accept the certainty of a lesser punishment, in situations where the probability of achieving an acquittal is tenuous, instead of risking the chance of a much greater punishment.

Thus, while expediency and administrative efficiency is the reason for the recent ascendancy of plea bargaining, acknowledging this motivation does nothing to diminish that there are at least four other good reasons the practice is invoked, and thus four other motivations / justifications for its use:

1. Plea bargaining by prosecutors is often motivated by the desire to secure convictions, for reasons of deterrence and desert, in situations where the case is not ‘air tight’.
2. Plea bargaining by defendants is motivated not by speed of sentencing, but instead, by the uncertain outcome of a trial and the relative severity of the sentence expected, if convicted.
3. Plea bargaining by defense attorneys is often motivated by the same concern as defendants, since their duty is, ultimately, to their client’s interests.
4. Plea bargaining saves victims of traumatic crimes, such as rape and child molestation, from having to relive those terrible experiences by allowing such cases to be prosecuted without going to trial.

With these indisputable alternative motivations at work, it becomes impossible to maintain that administrative efficiency is the sole motive or justification for plea bargaining, and more difficult to maintain that it is the primary motive or justification. In fact, some legal scholars believe that even if courts had unlimited resources to hold trials, seventy to eighty percent of all cases would still be settled via plea bargains.

Is Plea Bargaining Coercive?

Kenneth Kipnis has argued that plea bargaining is analogous to a contract made under duress, involving too much coercion for the choice to be considered voluntary. Since a contract made under duress is clearly unfair, and for that reason considered void, so too, he argues, is plea bargaining unfair if it can be shown to be done under duress.
Kipnis begins by pointing out that plea bargaining in the criminal law has many of the same features of the contract in commercial transactions. In both institutions offers are made and accepted, entitlements are given up and obtained, and the notion of an exchange, ideally a fair one, is present in both parties (Kipnis 1992, p. 245). He goes on to list examples of procedures in plea bargaining which he feels have analogues in contract law. His argument for drawing an analogy between plea bargaining and contract law is fine, up to this point. He runs into some trouble, however, when he attempts to draw an analogy between plea bargaining and contracts made under duress, where one party wrongfully compels another party to the terms of an agreement.

To elucidate his point, Kipnis cites the paradigm example: the agreement made at gun point. In such a scenario, a thief threatens to shoot someone unless they hand over their money. Facing a mortal threat, they readily agree to hand over their cash. But despite such consent, the rules of duress work to void the effects of such agreements (Ibid, p. 246). Kipnis believes that prosecutors can apply duress to defendants in similar ways. If not always with the threat of the death penalty, then by fear of the significantly harsher sentence a conviction rather than a plea bargain would receive. Both the gunman and the prosecutor require persons to make hard choices between a very certain smaller imposition and an uncertain, possible greater imposition. Since the gunman transaction is made under duress it is unfair. If plea bargaining is likewise made under duress then it follows that it too is unfair.

However, upon closer examination, the analogy breaks down. Kipnis’ argument is unsound because his analogy does not hold in two important regards. Kipnis would like to draw the analogy thusly:

**Coerced Money Transfer (Robbery)**
1. Party A’s life was threatened at gunpoint.
2. This has a coercive effect on party A’s decision.
3. Party A had no other reasonable option and assents under duress.

**Coerced Contract**
1. Party A acted wrongfully by applying coercion to party B.
2. This has a coercive effect on party A’s decision.
3. Party B had no other reasonable option and assents under duress.

**Plea Bargain**
1. Prosecutor threatens defendant with the possibility of a harsh or very harsh punishment.
2. This has a coercive effect on defendant’s decision.
3. Defendant has no other reasonable option and assents under duress.

There are two problems with such an analogy:
First, the U.S. legal system makes a distinction between coercion and wrongful or unlawful coercion, the latter being defined as, ‘The use or
threatened use of *unlawful* force’ (Black 2004, p. 542; emphasis added). For duress to be present wrongful or unlawful coercion must be present. The U.S. Supreme Court has several times reviewed the issue of whether plea bargaining, in general, constitutes *wrongful* or *unlawful* coercion (compromising the voluntariness of the plea). Each time they have ruled that it does not: *Shelton v. United States* (1958); *Brady v. United States* (1970); *Parker v. North Carolina* (1970); *North Carolina v. Alford* (1970); *Bordenkirker v. Hayes* (1978). Thus, we cannot maintain that the 2\textsuperscript{nd} part of the analogy applies to plea bargaining as an institution.

Second, it is never the case that the defendant has no other reasonable option. There is always the option to invoke one’s sixth amendment right to go to trial and challenge the prosecution to prove its case to all of twelve jurors beyond a reasonable doubt, after challenging the admissibility of evidence and witnesses prior to trial and the prosecution’s case at trial. You can assume there is no other reasonable option to giving the gunman your money (attempting to disarm them is an unreasonable option) (Black 2004, p. 542, emphasis added). If it is truly a coerced contract then a similar situation holds there – no reasonable option (for example, being forced to sign at gunpoint). However, if you refuse to plea bargain you will not get the threatened punishment unless you are proven guilty beyond a reasonable doubt in a trial by jury. In other words, another reasonable option exists. As long as this option exists plea bargaining is not analogous to the gunman example. Since trial by jury is a basic constitutional right, then we cannot maintain that the 3\textsuperscript{rd} part of the analogy applies to plea bargaining.

**The Case for Extra Punishment**

The logic of the plea bargain raises the following question: will a defendant who elects to stand trial, rather than plead guilty in expectation of a more lenient sentence, be punished more severely, if found to be guilty, for having exercised his or her constitutional right to trial by jury?

The rationale adopted by the U.S. Supreme Court, in *Brady v. U.S.* (1970), argued it was permissible to offer a lesser penalty in a plea because the state saved resources by not having a trial. That leaves the impression with some that ‘The defendant should be penalized for standing trial, but he should be rewarded for saving the court time (Alschuler 2003, p. 1). A criminal defendant often has a choice between a jury trial and a plea bargain. Since the plea bargain is an agreement to plead guilty in exchange for a more lenient sentence, then it follows, on this reasoning, that a conviction by trial must necessarily carry a harsher sentence. In *United States v. Wiley*, Chief Justice Campbell expressed just such a view:

> If then, a trial judge grants leniency in exchange for a plea of guilty, it follows, as the reverse side of the same coin, that he must necessarily forego leniency, generally speaking, where the defendant
stands trial and is found guilty (United States v. Wiley, 1960)

What opponents of plea bargaining are trying to propose with this argument is that if plea bargaining is an offer of leniency compared with the expected punishment if convicted by trial, then, judges must punish those found guilty at trial more severely to allow for the plea option to be more lenient. In other words, when judges pass sentence after trials they must forego considerations of leniency in sentencing in order to allow for plea bargains to be more lenient by comparison. This is sometimes construed to further entail that those sentenced after a trial are actually punished more severely than they deserve, since the state does not merely forgo leniency, but also acts punitively. As Timothy Lynch puts it, ‘Plea bargaining rests on the constitutional fiction that our government does not retaliate against individuals who wish to exercise their right to trial by jury’ (Lynch 2003, p. 26) To support such positions one can offer statistics suggesting trial convictions typically get longer sentences than those plea bargained counterparts as well as studies suggesting defendants convicted via trial are punished more severely than those convicted via plea bargain (McDonald 1982).

Taken prima facie, these can be seen to providing evidence in support of the argument that convicted defendants are treated more harshly and are thus penalized for going to trial. However, statistics and studies can be misleading. While it is true that in most cases defendants convicted by jury trial receive stronger sentences than those that are plea bargained, it does not follow that they are being punished more severely for going to trial. A National Institute of Justice report on plea bargaining found that:

Several studies suggest that tried defendants are punished more severely. Than pleaders but most of these studies have not controlled for variables which might account for both the fact that the case went to trial and the fact that it was sentenced more severely. A few studies which were able to control such confounding variables found that the mere fact of going to trial did not account for the sentence difference (ibid, 106).

In short, as political scientist Martin Hueman has pointed out, ‘Cases are different that go to trial than those that plea’ (Blankenship 2003, p. 7).

To clarify this point, consider the following hypothetical scenario: When a defendant is convicted by jury of crime X, he is subsequently sentenced by the presiding judge. While the judge does have some discretion in the sentencing process, he/she is ultimately limited to the parameters set by law. If the punishment for crime X is 10-15 years imprisonment, then the judge must operate within these parameters. While it may be true that the defendant was offered a sentence of 5 years in exchange for a guilty plea, and after electing to stand trial, was convicted and given the maximum of 15 years, it is not true that such a defendant was punished more severely for demanding a trial. Being
offered a lesser sentence to forego one’s right to trial does nothing to change the fact that one would have received, upon conviction, a sentence of 10-15 years anyway. The judge cannot sentence a defendant to the maximum of 15 years and then add on another 2-5 years for demanding a trial. It is also true that judges may give, and sometimes do give, if they deem it fitting, the most lenient sentence allowed for the crime. While it is true that this will surely be greater than the sentence of a plea bargain to a lesser charge, it is not the case that the judge was forced to give the stiffest sentence allowed for the crime in order to punish more severely than a plea would receive. Only under these circumstances could we say that the defendant was punished more severely for exercising their constitutional right to trial by jury.

Is Plea Bargaining a Systematically Unjust Way of Dispensing Punishment?

Perhaps the most debilitating charge against the practice of plea bargaining is that it perpetuates a legal system which continually doles out systematic injustice. We have all heard of disturbing cases where a person is convicted of a crime and found years later to be innocent, or where someone whom all the evidence points towards is released on a technicality or found not guilty by a jury. To be sure, these are troubling instances that remind us that no legal system is ideal. But, of course, these cases are aberrations, injustices that occasionally slip through the system, not injustices caused by the system itself. Plea bargaining, on the other hand, has been accused of dispensing not the occasional instance of aberrational injustice found in the trial by jury process, but instead, a more widespread systematic injustice. Kenneth Kipnis writes:

We can refer to these incorrect outcomes of a sound system of criminal justice as instances of aberrational injustice. In contrast, instances of systematic injustice are those that result in structural flaws in the criminal justice system. Here incorrect outcomes in the operations of the system are not the result of human error. Rather, the system itself is not well calculated to avoid injustice. What would be instances of aberrational injustice in a sound system are not aberrations in an unsound system: they are the standard result (Kipnis 1992, p. 253).

The general argument takes the following form: there are two types of defendants who plea bargain, those who are guilty of the crimes that they are charged with and those who are innocent. If an innocent person plea bargains, then they are punished for a crime that they did not commit. This is unjust. On the other hand, if a defendant who is guilty of a certain crime is punished for a less severe crime, via a plea bargain, then this is likewise unjust because they do not receive the punishment they deserve. It follows that the process of plea bargaining distributes systematic injustice because it punishes those who do not
deserve to be punished and does not punish those who do deserve punishment to the extent fitting their crime.

This could be taken as a powerful argument for the unjustness of plea bargaining. However, it is important to realize that the argument that the plea bargaining process distributes systematic injustice requires the demonstration of two dubious points: a) the theoretical point that retributivism (the theory that people should be punished only to the extent that they deserve) is the correct and sole moral justification for punishment, and thus retribution based strictly on desert should be the sole end of our criminal justice system; and b) the practical point that a retributivist theory of punishment could actually be implemented, given the resources it would require.

Consider, first, the theoretical problem: Here, a very brief review of utilitarian and retributivist justifications of punishment - which are certainly not exhaustive of theories of punishment - will serve to make my point. Simply put, a retributivist believes that punishment is justified by a system of desert; someone who commits a crime deserves to be punished. Furthermore, they deserve to be punished only to the extent warranted by the crime; no less and no more. It follows that on retributivist grounds plea bargaining is indeed systematically unjust. It either punishes those who do not deserve to be punished or it punishes those who do deserve to be punished less severely then they deserve. Kipnis, for example, assumes the retributivist framework:

For grades, like punishments, should be deserved. Justice in retribution, like justice in grading, does not require that the end result be acceptable to the parties... For bargains are out of place in contexts where persons are to receive what they deserve (Kipnis 1992, p. 253-254).

Utilitarians, on the other hand, believe that the fundamental principle of morality requires us to maximize the net utility for all in the long run. Thus, a utilitarian justification for punishment rests on the consequences of that punishment. On this model, we should not punish for the purpose of desert, but instead, for the purpose of deterrence.

It is now easy to see how a utilitarian and a retributivist could take opposite sides in the plea bargaining debate. Retributivists clearly must agree with Kipnis in holding that plea bargaining does out unjust punishment. On retributivist grounds, plea bargains never administer the deserved punishment; always punishing too little or too much. Utilitarians, however, can argue that plea bargaining is both more efficient and cost effective than a retributivist system, and that it offers more deterrence because many borderline cases that might not win at trial are prosecuted through plea bargaining. Furthermore, if we consider the large body of literature criticizing retributivism and promoting alternative models of punishment, it becomes obvious that providing the first necessary supporting argument would be an extremely difficult task.

Furthermore, even if we were to grant the debatable point that
Plea Bargaining’s Moral Controversies

retributivism is the proper justification for punishment, retributivists would still face the daunting practical problem of attempting to send every criminal charge to trial. In an ideal world, we would have the resources to accomplish this, but in reality, it is difficult to see how this could be done. Where would the personnel and funding come from? How would we avoid backlogging the legal system in a manner that would infringe on the constitutional right to a speedy trial?

Some states and cities have claimed to abolish plea bargaining without incurring a backlog of cases. When these kinds of claims are made it is important to look closely at exactly what is being banned. For example, in Alaska the Attorney General’s office announced a no plea bargaining policy for prosecutors. They could no longer dismiss, reduce, or alter charges solely to obtain a guilty plea. However, when prosecutorial plea bargaining ended, the courts immediately compensated by negotiating with defense council themselves, offering plea incentives through differential sentencing. By contrast, in El Paso, Texas, where they did truly ban plea bargaining in all of its forms, defendants realized that there was no reason for not going to trial. Therefore, they were reluctant to plead guilty and the El Paso court system soon developed serious backlog problems (McDonald 1982, p. 102-105).

In the end, for the argument that plea bargaining dispenses systematic injustice to be successful it must: a) justify retribution based strictly on desert as the sole end of our criminal justice system, and b) account for how a retributivist system of punishment could actually be implemented, given the resources it would seem to require. In the absence of compelling arguments to those effects the argument that plea bargaining dispenses systematic injustice is greatly diminished, if not diffused.

Conclusion

In sum, while administrative efficiency is the reason for plea bargaining’s ascendancy to wide spread use, it neither accounts for plea bargaining’s origin nor exhausts the motivations or justifications for its continued use. The argument that plea bargaining is analogous to a contract made under duress, is disanalogous and is therefore untenable. It is untrue that a defendant who elects to stand trial rather than plea bargain must be punished more severely, upon conviction, for exercising his or her constitutional right to trial. The argument that the plea bargaining process distributes systematic injustice is diffused by the realization that such a criticism requires the demonstration of two dubious points: a) the theoretical point that retributivism is the correct and sole moral justification for punishment, and thus retribution based strictly on desert should be the sole end of our criminal justice system, and b) the practical point that a retributivist theory of punishment can actually be implemented, given the resources it would require.

I recognize that the plea bargaining process is not without its problems. However, these need to be considered carefully: To the extent that defendants
plead guilty to crimes they did not commit because they fear a trial, that points to problems with our trial system rather than plea bargaining. Likewise, to the extent that defendants plead guilty to crimes they did not commit because of overbearing prosecutorial bargaining tactics (such as overcharging), that points to problems with the current procedural rules allowing such negotiating tactics - to contingent, rather than essential characteristics of plea bargaining - not to the right or process of making such a contract itself. Mere efficiency does not justify resorting to an immoral procedure. But, there are sufficient justifications for plea bargaining other than efficiency, and the flaws with the process of plea bargaining are procedural rather than essential; the practice, therefore, requires reform rather than abolition.

References
