AN ANTHOLOGY OF PHILOSOPHICAL STUDIES
VOLUME 5

Editor
Patricia Hanna
University of Utah
USA

Editorial Board

Gerald Doppelt
University of California, San Diego
USA

Carol Nicholson
Rider University
USA

Donald Poochigian
University of North Dakota
USA

T. Ann Scholl
United Arab Emirates University
UAE
**Board of Reviewers**

Daniel Considine  
Metropolitan State College in Denver, Colorado  
USA  
Katherine Cooklin  
Slippery Rock University of Pennsylvania  
USA  
Gary Fuller  
Central Michigan University  
USA  
Keith Green  
East Tennessee State University  
USA  
Dimitria Electra Gratzia  
University of Akron  
USA  
Crista Lebens  
University of Wisconsin-Whitewater  
USA  
Michael J. Matthis  
Lamar University  
USA  
Mark McEvoy  
Hofstra University  
USA  
Joseph Mintoff  
University of Newcastle  
Australia  
Chris Onof  
Birkbeck College  
UK  
Elizabeth Schiltz  
College of Wooster  
USA  
John Thompson,  
Christopher Newport University  
USA  
Andrew Ward  
University of York  
UK
First Published in Athens, Greece by the Athens Institute for Education and Research.

ISBN: 978-960-9549-24-0
All rights reserved. No part of this publication may be reproduced, stored, retrieved system, or transmitted, in any form or by any means, without the written permission of the publisher, nor be otherwise circulated in any form of binding or cover.

Printed and bound in Athens, Greece by Theta Co.
Printed and bound in Athens, Greece by ATINER SA

8 Valaoritou Street, Kolonaki
10671 Athens, Greece
www.atiner.gr

©Copyright 2011 by the Athens Institute for Education and Research. The individual essays remain the intellectual properties of the contributors.
# Table of Contents

## List of Contributors

1. **Introduction**  
   Patricia Hanna  

2. **Knowing the Sophist: Encountering the Law**  
   Hagit Aldema  

3. **The Transvaluation of Representation**  
   Kieran Anthony Cashell  

4. **Euripides and Plato on Autochthony**  
   Jaroslav Danes  

5. **The Myth of Gyges and the Possibility of Altruism**  
   Chrysoula Gitsoulis  

6. **Aristotle’s Theory of Responsibility and Situationism**  
   Panos Kapetanakis  

7. **Geometry without Space: Ancient Greek Mathematical Thought and Contemporary Consequences**  
   Duane Lacey  

8. **A Comparative Study of the Role of Harmony in the Thoughts of Confucius and Plato on the Nurturing of Culture through Music**  
   Mei-Yen Lee  

9. **On the Issue of the Word Φιλοσοφία in its Original Sense**  
   Vasily Vasilievitch Markhinin  

10. **Friendship and the Common Life: Aristotle’s Contribution for a Modern Utilitarian Society**  
    Philip Matthews  

11. **Reconceiving Philosophy as Bodily Discipline: Clues from Master Zhu**  
    John M. Thompson  

   *Modern, 19th and 20th Century*

12. **Descartes, Luther, and the Fifth Lateran Council**  
    Adermi Artis  

13. **Schopenhauer's Critique of Kant and the Ambiguous Role of Kantian Inclinations**  
    Ryan Beaton  

14. **The Logic of Heidegger's Rectorship Address**  
    Richard Hearn  

15. **The Problem of the Censor**  
    Elizabeth A. Murray  

16. **An Adverbial Interpretation of Pragmatism**  
    Carol Nicholson  

17. **Merleau-Ponty's Doctrine of Objects**  
    Henry Pietersma
| 18. | The Necessity of the “Necessary Connection” in Kant and Hume: Reflections on an Old Challenge to Metaphysics |
|      | Irmgard Scherer |
| 19. | Rorty and the Linguistic Turn |
|      | Timo Vuorio |

**Part 2: Metaphysics, Logic and Philosophy of Language**

| 20. | Compatibilism and the Folk Psychology of Free Will |
|      | Gregg Caruso |
| 21. | Personal Identity and Stories |
|      | Gary Fuller |
| 22. | Why there is 'No Such Thing as Language': Chomsky's Mistake |
|      | Patricia Hanna |
| 23. | Bound Yet Free: An Elucidation and Defense of Davidson's Principle |
|      | Michael T. Michaelakis |
| 24. | Ontology that Matters: Binding Relations |
|      | Joseph Naimo |
| 25. | On the Nominal Constitution of Mathematics |
|      | Donald Poochigian |
| 26. | A Physicalist and Explanatory Concept of Propositional Attitudes |
|      | Krzysztof Swiatek |
| 27. | Structural Norms in Light of Neuroscience |
|      | Elżbieta Szymanska-Switk |

**Part 3: Ethics and Value Theory**

| 28. | The Role of Sociological Assumptions in the Construction of Ethical Theories |
|      | Marcos G. Breuer |
| 29. | Thoughts by the Sea, Reflections on Happiness |
|      | Louis Colombo |
| 30. | Intergenerational Rights: A Philosophical Examination |
|      | Makoto Usami |
List of Contributors

**Hagit Aldema** is an adjunct lecturer at Tel-Aviv University, Israel. Her main research interests include Ancient Philosophy, Continental Philosophy and Psychoanalysis.

**Aderemi Artis** is Assistant Professor of Philosophy at the University of Michigan-Flint. He works in early modern European philosophy, more specifically on the impact of scientific and religious ideas on canonical philosophical themes and figures. Other interests include metaphysics and ethics.

**Ryan Beaton** received his PhD in Philosophy from the University of Toronto in March 2011. His dissertation examines the opposition of Kant, Schopenhauer and Nietzsche to eudaimonism and utilitarianism, with a particular focus on Nietzsche’s critique of Spencer and Mill. He recently co-authored the entry on *eudaimonism* for the *International Encyclopedia of Ethics*, and is currently studying law at Harvard Law School, where his interests include especially critical legal studies and legal history of the 19th and 20th centuries.

**Marcos Breuer** is a Post-Doctoral Research Fellow at the Department of Philosophy and History of Science, National University of Athens, Greece. His research centers on the construction of a multiplex model of moral motivation as a basis for universalistic normative theorizing. His doctoral thesis analyzes Mackie, Hare and Rawls’ moral-psychological assumptions; it has been published in German by LIT-Verlag (*Soziologie in der Ethik: Handlungs-theoretische Fundamente der Moralphilosophie*). Other research interests include Social Theory and Meta-Ethics.

**Gregg Caruso** is Assistant Professor of Philosophy and Chair of the Humanities Department at Corning Community College, SUNY, New York, USA. His research interests include philosophy of mind and metaphysics, with a particular interest in consciousness and free will. His most recent work focuses on the phenomenology of agency and the folk psychology of free will. His broader work engages issues at the intersection of the behavioral, cognitive, and neurosciences. His papers have appeared in such journals as *Southwest Philosophy Review* and *Kriterion: Journal of Philosophy*.

**Kieran Cashell** lectures in the Limerick Institute of Technology and is author of *Aftershock: the Ethics of Contemporary Transgressive Art* (I.B. Tauris: 2009). His principal area of investigation is the problem of representation which traverses the fields of Epistemology, Philosophy of Mind, Language and Art. An interest in the uneasy relationship between artistic value and moral value has also resulted in an extended inquiry into the intersection of Ethics and Aesthetics.

**Louis Colombo** is an Assistant Professor of Philosophy at Bethune-Cookman University, a historically black university in Daytona Beach, Florida, USA. His research interests include Modern Philosophy, Hegel, and Critical Theory.
Jaroslav Danes is an Assistant Professor at the Institute of Philosophy and Social Sciences of the Faculty of Philosophy University of Hradec Kralove. He has written articles and books in Czech and English. His main research interest lies in the field of History of Political Thought and Political Drama.

Gary Fuller is a Professor of Philosophy at Central Michigan University. He has published on a number of topics, including simulation theory, personal identity, and empty names.

Chrysoula Gitsoulis is an Adjunct Assistant Professor in Philosophy for the City and State Universities of New York, and Stevens Institute of Technology in Hoboken, New Jersey. She has presented paper at over 50 philosophy conferences, and has published several articles on a variety of topics. Her research interests focus on Ancient Greek Philosophy, Moral and Political Philosophy, and the Philosophy of Language (especially Wittgenstein).

Patricia Hanna is Professor of Philosophy and Professor of Linguistics at the University of Utah, Salt Lake City, UT, USA. She has published on philosophy of language, ethics, children's rights, philosophy of mind, belief and reference, ontology, the realism/relativism debate, and Wittgenstein. She is the co-author, with Bernard Harrison, of *Word and World: Practice and the Foundations of Language*. Her current work focuses on Chomskyan theoretical linguistics from a Wittgensteinian perspective.

Richard Hearn M.D., PhD., J.D. graduated magna cum laude from Georgetown Law Center in Washington D.C. and currently is a partner in the Idaho law firm of Racine Olson, Nye Budge & Bailey. Following graduation from law school in 1991, he was a visiting scholar at the Center for Ethics in Science located in Tubingen Germany. He retired from the practice of medicine in 2003. This spring Dr Hearn will present papers at both the Annual Meeting of the Heidegger Circle to be held at Marquette University and the 2011 meeting of the Martin-Heidegger-Forschungsgruppe to be held in Messkirch Germany. His research interest includes all aspects of Heidegger scholarship.

Panos Kapetanakis is a Ph.D student of Philosophy at the University of Edinburgh, UK. His main research interests lie in ethics, moral psychology and political philosophy. His other interests include political, economic and social theory.

Duane J. Lacey is an Assistant Professor of Philosophy at United Arab Emirates University, Al Ain, UAE. He earned all of his degrees at The New School for Social Research; his PhD Dissertation was on “Euclid’s Taxonomy of Irrationals.” He has worked as a Visiting Scholar at the University of Notre Dame, a Visiting Lecturer at Borganes University in Iceland, a faculty member (Tutor) at St. John’s College in Annapolis, MD, and a Visiting Professor with New Gelati Academy and Grigol Robakidze University in Tbilisi, Republic of Georgia. His research interests and previous research include: The History of Philosophy, Political Philosophy, The Philosophy of War, Ancient Science and Mathematics, Greek Philosophy, and The Graeco-Arabic Translation Movement.

Mei-Yen Lee is a Professor in the Department of Chinese Language and Literature at National Pingtung University of Education, Taiwan, Republic of China. Her main research interest lies in the field of Chinese musical aesthetics, specializing in ancient Chinese seven-stringed Lute which is known in Chinese as the *ch‘in* or *guqin*. She has written 5 books and more than 80
pieces of academic papers. Her recent authored work is *Ch’in-Tao: Robert Hans van Gulik and the Chinese Lute*. (Hong Kong: Jao Tsung-I Petite Ecole, the University of Hong Kong, 2010).

**Vasily Vassilievitch Markhinin** is Professor of Philosophy at the State University in Surgut, Russia. Research interests include ethnosophological researches, theories of ethnosoph, mythological and religious world-view, the philosophy of science issues, the problem of specific character of philosophy as a type of cognition and world outlook, Ancient Greek Philosophy. He has published in various journals in both Russian and English.

**Philip Matthews** was previously Director of the Bioethics Teaching and Research Unit in the School of Public Health, Curtin University and editor of *Faith and Freedom: a journal of social ethics*. He is currently undergraduate coordinator for the philosophy program in the School of Philosophy and Theology, University of Notre Dame, Fremantle.

**Michaelakis T. Michael** is Assistant Professor of the History and Philosophy of Science at Underwood International College, Yonsei University, Seoul. He works primarily in Philosophy of Action and Philosophy of Psychoanalysis. His papers have appeared in journals such as *Studies in the History and Philosophy of Science* and *Cheolhak*.

**Elizabeth Murray** is a Professor of Philosophy at Loyola Marymount University in Los Angeles, California, USA. She works on the thought of Bernard Lonergan and in the area of existential phenomenology. She is co-editor with Mark Morelli of *The Lonergan Reader* (Toronto: University of Toronto Press, 1997). She has published on topics dealing with the thought of Kierkegaard, Nietzsche, Husserl, Sartre, and Lonergan. She is the author of *Anxiety: A Study of the Affectivity of Moral Consciousness* (University Press of America, 1986).

**Joseph Naimo** lectures in Philosophy and Ethics at the University of Notre Dame Australia, Fremantle. Active areas of research include Consciousness studies, Metaphysics and Ethics. Most recent publication - Naimo, J. (2009) *The Primacy of Consciousness: A Triple Aspect Ontology*. Areas of interest: philosophy of science; and comparative religious studies.

**Carol Nicholson** is Professor of Philosophy at Rider University in New Jersey, USA. Her research interests include American pragmatism and philosophy of education, and she has published articles in such journals as *Transactions of the C.S. Peirce Society, International Journal of the Humanities, Educational Theory*, and *Philosophy Now*. Her most recent paper, “Education and the Pragmatic Temperament,” will be published in the *Cambridge Companion to Pragmatism*.

**Henry Pietersma** is Emeritus Professor of Philosophy at the University of Toronto, CA. He has published numer articles on phenomenology. He edited *Merlewu-Ponty: Critical Essays* (Center for Advanced Research in Phenomenology), and is the author of *Phenomenological Epistemology* (Oxford University Press).

**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. Illustrating his somewhat eclectic concerns,

**Irmgard Scherer** is Associate Professor of Philosophy (Emerita) at Loyola University in Baltimore, Maryland, USA. Her work is in the 18th century history of ideas, specializing in Kant’s aesthetics and the problem of metaphysics. She has written many articles and a monograph, *The Crisis of Judgment in Kant’s Three Critiques (In Search of a Science of Aesthetics)*, which spans Kant’s critical corpus in science, ethics and art from the standpoint of human judgment.

**Krzysztof Swiatek** is a Philosophy Instructor at Grant MacEwan University in Edmonton, Alberta, Canada. His main research interest lies in the field of philosophy of mind. Other research interests include political philosophy, hermeneutics, philosophy of psychology and epistemology. He published papers on narrow content and self-knowledge as well as L. Kolakowski’s philosophy of ideology, L. Nowak’s political philosophy and Ricoeur’s hermeneutics.

**Elzbieta Szymanska-Swiatek** teaches philosophy at Grant MacEwan University in Edmonton, Alberta, Canada. Her research interests include continental philosophy, philosophy of mind and moral psychology. She published on Bergson and Husserl. Her current research focuses on neurophysiological and experimental foundations of moral judgments.

**John M. Thompson** is Associate Professor of Philosophy and Religious Studies at Christopher Newport University (in Newport News, Virginia). He has broad interests in both Asian and Western philosophies and religions, and has published numerous articles and book reviews. His most recent book, *Understanding Prajna: Sengzhao’s ‘Wild Words’ and the Search for Wisdom*, critically examines early Chinese Buddhist views of enlightenment, and was published by Peter Lang in 2008.


**Timo Vuorio** is Teacher and Researcher of Philosophy at School of Social Sciences and Humanities, University of Tampere, Finland. He works on metaphilosophy and on the issues of ‘first philosophy’ generally, but his special interest is Richard Rorty’s philosophical legacy and its applications, especially to philosophy of technology. His philosophical orientation is neopragmatist. His papers have appeared in such journals as *Acta Philosophica Fennica* and *Human Affairs*.
CHAPTER ONE

Introduction

Patricia Hanna

This volume is a collection of papers selected from those presented at the 5th International Conference on Philosophy sponsored by the Athens Institute for Research and Education (ATINER), held in Athens, Greece at the St. George Lycabettus Hotel, June 2010.

Held annually, this conference provides a singular opportunity for philosophers from all over the world to meet and share ideas with the aim of expanding our understanding of our discipline. Over the course of the conference, sixty papers were presented. The twenty-eight papers (one of which was originally presented in 2009) in this volume were selected for inclusion after a process of review by at least two of the editors and reviewers. We have organized the volume along traditional lines. This should not, however, mislead a reader into supposing that the topics or approach to problems fall neatly into traditional categories.

The selection of papers chosen for inclusion gives 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.
PART 1
History of Philosophy
Ancient: Greek and Eastern
C H A P T E R TWO

Knowing the Sophist: Encountering the Law

Hagit Aldema

The focus of the following essay is the moment of the interlocutors' act of parricide in Plato's dialogue *Sophist* (Sophist 241d; 355). This moment will be examined and interpreted through a psychoanalytic reading of the myth of parricide as described by Sigmund Freud in *Totem and Taboo* and later developed by Jacques Lacan. My interpretation of parricide is supported by psychoanalytic thought which deems it a myth related to the Subject’s birth. Psychoanalysis indeed demonstrates that the mythical birth of the Subject is conditioned by the act of parricide.

The act of parricide will be discussed in this paper in the context in which it occurs in Plato's *Sophist*; the interlocutors' rethinking of the dichotomy of being and not-being and their agreement 'that after a fashion not-being is and on the other hand in a sense being is not'. This conclusion is a violation of the philosophical taboo formulated by Parmenides, the 'father' of their discourse, as he is referred to by the interlocutors.

Since the main aim of Plato's *Sophist* is to know the sophist's location in 'the darkness of not-being' (as it is regarded by the interlocutors), the violation of Parmenides' taboo seems unavoidable. For it is the purpose of the investigation which already anticipates the parricide.

What is the relation between the interlocutors and Parmenides' taboo? Is it possible, really, to violate the taboo of the father? Doesn't an investigation of the sophist undermine the possibility of violating the taboo? Is it unnecessary to say that the sophist is almost primordially defined as the stepson of Parmenides, the one who speaks of not-being as being? And if such is the case, what does the violation of Parmenides' taboo turn his legitimate sons into? All these questions relate to and merely orbit around the primary inquiry of the present paper; the possibility or impossibility of distinguishing between the sophist and the philosopher.

Taboo and Desire

The Freudian myth describes the a-temporal phase of humanity that precedes socialization, wherein one omnipotent father controls the group and owns, by way of force, all the women in it, forbidding them from his sons. This prohibition of the women causes the sons to murder the patriarch in the hopes of gaining access to the women. What Freud offers by way of this myth is that the freedom the sons wished to achieve by the act of parricide, served merely
to engender strife and ultimately murder amongst the sons. In order to stop this war of 'all against all', the sons establish the taboo of incest, which allows them a degree of enjoyment, albeit not absolute as their father had, but one that enables them to exist both as a society and as individuals.

'The new organization would have collapsed in a struggle of all against all, for none of them was of such over-mastering strength as to be able to take on his father's part with success. Thus the brothers had no alternative, if they were to live together but [...] to institute the law against incest, by which they all alike renounced the women whom they desired and who had been their chief motive for dispatching their father' (Freud 2001:167).

Meaning, the act of parricide intended to allow the sons the freedom enjoyed by their father, leads them, in effect, to establish the father's prohibition by making an enforceable law of it—a taboo. In an attempt to regulate their relationship—destroyed by their shared desire to obtain the prohibited women—the sons opt to make their own prohibition law, their own taboo. Indeed, the taboo reveals the structure of a force choice: the sons must choose between fighting each other to the death for the privilege of enjoying the women or establish a law which, while forbidding them the enjoyment of obtaining all the women as their father did, will allow them to individually enjoy certain women defined by the law as attainable.

Drawing on Freud and merging it with linguistics, for Lacan the law of the father is fundamentally a linguistic entity. It is the law of the signifier that is both the cause of desire and that which poses a limit to it. As is pointed out by Lacan, the Subject’s modes of desire are determined by the law of the signifier.

As Lacan refers to the myth of parricide, particular emphasis is placed on its structure and therein the structural bond between law and desire. Indeed, the myth of primal parricide uncovers an impossible relation between the Subject and the law, an impossibility bound to the paternal function. The father's prohibition, imposed by force, once undertaken becomes a prohibition the sons enforce upon themselves in the form of a law.

This structure reveals that desire cannot reach its ultimate satisfaction since it is always-already (according to the a-temporality of any myth as such) mediated by law. Lacan’s thinking stresses that the relation between the law and desire is a dialectical one; 'desire is the flip side of the law' (Lacan 2006: 665). Law simultaneously creates desire by introducing prohibition and imposes limits to desire. Given that the primary constituent of desire is the desire to transgress, it seems clear that prohibition, a limitation, is necessary for the transgression to take place—as that from which to transgress. As evidenced in the myth of primal parricide, the law is structurally that which causes the Subject’s desire yet inserts a built-in boundary to it. This barrier, as psychoanalysis would have it, is critical to the formation of desire. A desire born, in other words, from the mere presence of the boundary.

In following psychoanalysis' articulation of parricide it is revealed that the
primary encounter with the law exemplifies an impossible relation between it and the Subject. This impossibility is a structural one and therefore can never be altered into a reciprocal content relation between the Subject and the law. Being the catalyst to the subject’s desire and that which disables it, the law, in essence, commands the Subject's modes of desire. Desire which, contrary to need, is never fully satisfied, and as such creating an encounter relation between the Subject and the law.

The psychoanalytic context I present here and therein the relation between the Subject and the law may help us shed light on the task at hand—that of 'knowing the sophist'—which, in keeping with the psychoanalytic perspective, is concerned with the desire to know rather than with the achievement of an actual knowledge grounded in the fundamental 'incompatibility between desire and speech' (Lacan 2006: 275)

The desire to know the sophist, as it is conceptualized in the psychoanalytic mindset, relates the Subject (the interlocutor) to its object of desire (the sophist), and this object of desire is accessible only as an object of lack. The desire to know, much like desire in general, relies on the relation between the Subject's demand for pleasure (knowledge in Plato's universe) and the matter out of which this pleasure is made; namely real loss that is articulated by the signifier.

Philosopher versus Sophist?

Several readings of the Sophist claim that it is composed of two 'external' independent texts (Sophist 216a-236d, 264b-268d) in between which lies the 'middle section' that is the dialogue (ibid, 236d-264b). The 'middle section', as it were, touches on questions of being, not-being and the status of truth and falsehood, while the 'external' sections that wrap around it discuss the definitions of the sophist. Based on its structure, various interpreters of the Sophist have claimed that the two external parts were added by Plato at a later stage and form an artificial addition to the central crux. To refute this claim, we may wish to read the Sophist as a dialogue between the Stranger and Theaetetus in which numerous 'evidences', so to speak, of the text's greater, overall, aim to know the sophist (to define him) come to light (ibid, 236d, 239c, 240d, 260-261a, 264e). This 'evidence' refers to the matters raised by the interlocutors in the middle section of the dialogue in order to solve the difficulties that come to light as they try in vain to define the sophist. In brief, the question of the sophist's being is the central issue of this dialogue and the efforts to answer this question inherently touch on all aspects of the piece.

If we are to place the different issues raised throughout the dialogue under the umbrella question of the sophist's being, we cannot but acknowledge that the demand to define the sophist's being is invariably bound to that of the philosopher's being. The sophist cannot be defined independently from the philosopher and vice versa. This symmetry of one's dependence upon the other deems the questions of being and not-being (raised in the 'middle section')
equally dependent upon each other.

The interlocutors' knowledge of the sophist appears in the form of seven definitions which are attributed to the latter's discursive craft. Their goal is to provide a distinction between the sophist and the philosopher. I shall focus at this time on the sixth definition (ibid, 230-231b; 315-317); the definition which, I believe, leads the interlocutors to parricide.

The sixth definition which draws on the Socratic method of inquiry, fails to provide the desired distinction between the sophist and the philosopher. Rather than sketching a distinction between them, the sixth definition discloses a similarity between the two by way of crowning Socrates the ideal sophist. This definition of the sophist formulates in a more intense manner the necessity of having the sophist's definition in order to provide a distinction between him and the philosopher but, simultaneously, it discloses a crucial impossibility: a fundamental aporia.

As a means of solving this aporia (which doesn't at all considered as such), following a moment of hesitation after granting the sophist 'too high a meed of honor' the Stranger nevertheless decides that 'the strife will not be about petty discriminations' (ibid, 231; 317). Meaning, the interlocutors are not turned away by this problem, to say the least. Rather than abandoning their way of investigation which had thus far lead them to disclose similarity rather than to achieve a difference between the sophist and the philosopher, rigorously continue with their dialectic method of investigation (referred to here by one of its facets, diaerisis). The question the reader becomes eager to ask then is, why; why continue in the same vein of investigation?

The acceptance of the sixth definition, which derogates Socrates' elenchus to the position of mere sophism, lies in the Stranger's ethical decision which is, in effect, indifferent to ethics. The Stranger's method of investigation aims at deciphering what is and what is not related in all the arts for achieving knowledge. In the Stranger's words his method: 'is indifferent if the one benefits us little, the other greatly by its purifying. [...] and therefore it honors them all equally and does not in making comparisons think one more ridiculous than another' (ibid, 227b; 303-305). From this citation we may deduce that the Stranger's method does not allow him to reconsider, much less so to abandon, the achieved definition of the sophist which corresponds with the demand to remain indifferent to the sophist's honor even at the expense of identify him with Socrates.¹

But if we agree that the skopos (the purpose) of Plato's Sophist is to define the sophist for the purpose of distinguishing him from the philosopher, then we have just demonstrated how the dialogue's sixth definition, due to the Stranger's ethical neutrality or, we may say, his rising above ethics, decidedly fails to offer this distinction. Since the desire to distinguish between the sophist and the philosopher is a Socratic desire. And what is the Socratic desire if not directed at demonstrating, through and through, the ethical superiority of the philosopher, using the image of the sophist to do so?

¹See Gonzalez 2000: 162-168
What we have here is indeed the legacy of the Socratic desire to validate the superiority of the philosopher in relation to the sophist but in the form of its denial. In other words, the only thing that may lead to the desired distinction between the sophist and the philosopher, to the Stranger's successful investigation as it were is the Socratic axiom of philosopher's superiority. Only that this axiom is precisely that which the Stranger's method of investigation resists. This is, ultimately, the cause for which the investigation ends in failure (despite the arrogant tone by which it ends). Indeed, by the end of the dialogue the Stranger poses a question before Theaetetus, asking whether the man upon whom the investigation is performed is philosopher or sophist, sophon or sophistikon. In reply to his inquiry Theaetetus answers that they may not call him wise because by their divisions he is ignorant, he lacks knowledge, and therefore he is the 'actual sophist' they have been looking for (ibid, 268 b-c; 457, 459). Like the sophist, Socrates claims for ignorance, this is his ethical position towards knowledge. If so, the end of this investigation succeeds at most in illustrating that Socrates is a sophist. It does not, however, define (know) the sophist.

Encountering the Law

There is a short distance it seems, between crowning Socrates the ideal sophist and committing parricide in order to know the sophist. There is not enough words to analyze here the division that precedes parricide but only to note that if we follow the 'ethical indifference' of the Stranger's methodological decision, we ought to accept that the Stranger discriminates (i.e. desires to discriminate) on the basis of like from like rather than on the basis of better from worse. The division created between two forms of the image making art; 'the likeness-making and the fantastic' (ibid, 236c; 337) does not honor them alike. The fantastic images which the Stranger tends to attribute to the sophist are misleading images; they look like they are reflective of the original yet in actuality they are not.

However, following the Stranger's method which locates the sophist at the segment of 'fantastic art maker' opens a crucial and irreducible gap into the dichotomy between being and not-being. Indeed, it leads the Stranger to admit 'that falsehood really exists and in saying this [means] to be involved in contradiction' (ibid, 237; 337), namely, to attach not-being to being, to encounter, in essence, Parmenides' taboo.

Let us look at Parmenides' prohibition precisely as it is worded in the dialogue:

'Never let this thought prevail, saith he, that not-being is;
But keep your mind from this way of investigation' (ibid 237; 339)

---

1For more about the Stranger's methodological decision see Rosen 1983: 120-121, 135-137, 326-328
The interlocutors are required to violate Parmenides' taboo in order to make of the sophist what they wish to make of him; an artist of deception. Indeed, deeming him such is made possible only by way of inquiring into the ontological status of falsehood. 'The darkness of not-being' in which the philosopher always-already locates the sophist is a darkness that, according to a literal reading of Parmenides' taboo, is forbidden to think of. In the context of this prohibition, it would not be too great a stretch to claim that the dialogue concerning the sophist is from the start a dialogue which's object is forbidden from thought. In aligning with this claim, it would seem that the sophist, by way of his art of deception causes the interlocutors to lead their investigation in the direction that Parmenides commands should not prevail.

To which object can we attribute the speech of 'not-being'? Parmenides' taboo forbids answering this question but demands that 'not-being' should never be ascribed to 'being'. This complete distinction between being and not-being leads interlocutors to the conclusion that, as Parmenides orders, being and not-being shall never be linked. However, in inquiring into the nature of not-being—into the essence of 'things that are not'—the very inquiry exposes, at once, the oneness and plurality attached to not-being. For that the interlocutors determine that not-being is 'inconceivable, inexpressible, unspeakable, irrational' (ibid, 238c; 343). Only that this conclusion helps them to unravel the tie between not-being and is, it helps the interlocutor's to refer to not-being as an object that they express and speak about - not-being that is. This philosophical inconsistency comes to its resolve once the interlocutors allow themselves to think about being that is-not and not-being that is. Such thought process allegedly refutes and overrides Parmenides' taboo which is treated by them as a thesis and henceforth serves as a cornerstone for their further investigation.

'Str.: It seems to me that Parmenides and all who ever undertook a critical definition of the number and nature of realities have talked to us rather carelessly[...]every one of them seems to tell us a story, as if we were children' (ibid 242c; 357, 359).

As these children ('as if we were children') grow older it would seem from the citation above that they too will perpetuate the act of telling a 'careless story'.

Psychoanalysis concerns itself with a rather different structure as it pertains to the paternal taboo. For psychoanalysis, the taboo is hardly a thesis and therefore cannot be contradicted as a thesis could. Indeed, psychoanalysis' perception of the taboo is that the very command not to think (in this case, not to think that not-being is) already testifies to the desire to think the forbidden thought. This claim is made clear by way of Lacan's distinction between philosophy and psychoanalysis: 'the Freudian world is not a world of things, it is not a world of being, it's a world of desire as such' (Lacan in Badiou 2008: 228). In this context, the taboo represents, in its very existence, a function of desire; a way by which the Subject unconsciously regulates his modes of
desire. In Freud's words:

'Since taboos are mainly expressed in prohibitions, the underlying presence of a positive current of desire may occur to us as something quite obvious and calling for no lengthy proofs based on the analogy of the neuroses. For, after all, there is no need to prohibit something that no one desires to do, and a thing that is forbidden with the greatest emphasis must be a thing that is desired' (Freud 2001: 80-81).

The taboo, as Freud sees it, already assumes the Subject's transgression of it. Indeed, the Parmenidean taboo testifies to philosophy's desire to transgress it. When it is said 'Never let this thought prevail that not-being is', the informer of the law commands not to think that not-being is. Hence, he formulates the mode of thought he expects the addressees to follow: 'keep your mind from this way of investigation'. It is of importance here to pay close attention to the literal meaning of the quote. The addressees are not requested not to think of not-being as being, but to 'keep [their] mind' from this thought; to abstain from such thought altogether. Yet how far is the distance between not thinking of not-being as being and not thinking of this question at all? When is it far enough? Or close enough? As these questions indicate, the addressees must know what they are expected to refrain from, yet in order to know that they first need to dwell on the thought that 'not-being is'. Simply put, the response to the law necessarily leads to its transgression because it cannot be dealt with without already transgressing it; without the attempt to think what the law demands should not be thought.

The sophist, as the desired object of this dialogue, leads the interlocutors to parricide: 'you see, at any rate, how by this interchange of words the many-headed sophist has once more forced us against our will to admit that not-being exists in a way' (Sophist 240; 351). Yet the violation takes place as a response to the father's law—it is a means by which to prevent the sophist from claiming to be an equal addressee of the taboo. In their attempts to define (know) the sophist the interlocutors are required to act against Parmenides' law, yet are simultaneously required to remain subjected to it; to continue their investigation and maintain the semblance of being which they attribute to the sophist's deceptive art in order to exclude him from the realm of dialectic discourse.

It is the law of the father that is after all rooted at the base of their thought despite their assertion to violate it. The investigation proceeds according to the rules of the dialectic discourse—that which, by definition, is directed towards being—the discourse that is controlled by Parmenides as well as by their antecedes fathers.

The relation between responding to the law of the father and transgressing it becomes clear in light of the above claim. As Freud structures it in Totem and Taboo, the act of parricide unveils the impossibility of truly transgressing the law of the father. Since the law already bespeaks the desire to transgress it,
desire and law cannot be separated: 'desire is the flip side of the law' (Lacan 2006: 665). As we can see, the necessity to define the ontological status of not-being (toward which the interlocutors are thrown in their pursuit to know the sophist) is the very desire that unavoidably turns the father’s prohibition into a full fledged law. We may like to say that the act of parricide is nothing but a response to the father’s prohibition that takes on the form of excluding the sophist from the dialectic discourse.

We can see that the transgression of the law is nothing but a lure that is placed within the law itself. That occurs here in relating to the sophist’s negation of the existence of falsehood, a negation which can easily collaborate with Parmenides. To think that falsehood exists is to think that ‘not-being is’ and to neglect the existence of falsehood, like the sophist, is paradoxically to become the legitimate son of Parmenides. The distinction that results from Parmenides’ taboo appears to be a distinction that the sophist attributes to himself. It is the sophist who argues for an absolute distinction between being and not-being in order to claim that there isn’t an ‘existence of falsehood’ or that there isn’t a possibility that discourse can relate to not-being. And so it is precisely the preconception of the sophist as deceptive; making false statements and fantastic art, that plunges the inquiry into an abyss of philosophical puzzlement. Indeed, it is at this issue that the sophist, by way of his negation of the existence of falsehood, escapes the place the inquirers desire to locate him in (ibid, 236e-237a; 241 a-b). This issue is recalled at a later part of the dialogue.

‘Str.: But we said that the sophist had taken refuge in this region and had absolutely denied the existence of falsehood: for he said that not-being could be neither conceived nor uttered, since not-being did not in any way participate in being [...] But now not-being has been found to partake of being, and so, perhaps, he would no longer keep up the fight in this direction; but he might say that some ideas partake of not-being and some do not, and that speech and opinion are among those which do not; and he would therefore again contend that the image-making and fantastic art, in which we placed him, has absolutely no existence, since opinion and speech have no participation in not-being; for falsehood cannot possibly exist unless such participation takes place’ (ibid, 260c- 260e; 429).

When we claim that the sophist is an artist of deception; that he states falsehood, the sophist himself claims against us that we contradict ourselves. By claiming that falsehood exists, the sophists replies, we attribute not-being to being (ibid, 241 a-b; 353). To reject the possibility of falsehood means to eliminate any basis for the distinction between true or false speech and therein to eliminate any possibility of contradiction. It is in this way that the sophist's
imagined claims against inquirers' contradiction (ibid, 236e-237a; 337, 238d-239a; 345, 239c-d; 347,241d-e; 355), in effect, trap the sophist himself in a contradiction. Indeed the sophist find himself unable to contradict the philosopher's contradiction of himself. The only way to break this cycle is, again, to deny the existence of falsehood, which means to come back to the shelter of Parmenides' patronage from which they had never escaped. The sophist's denial of falsehood alongside the claim that 'opinion and speech do not participate in not-being' allows the paradoxical possibility of the sophist claiming that there is no principled difference between his discourse and that of the dialectician.

The final and concluding definition of the Sophist leads us readers to believe that the mission of knowing the sophist has been completed and that, in completing the process of knowing the sophist, he has been effectively placed outside the dialectic discourse. Furthermore, this final definition offers the reader the impression that the interlocutors had entirely overcome the difficulties placed before them at the onset of the dialogue by the sophist as well as by Parmenides' taboo. Only that the law of the father cannot ever be fully negated or violated. The Subject remains subordinate to it precisely because the violation has occurred. The subject's subordination to the law is perhaps best articulated by his attempts to understand the semblance of being in the sophist's discourse. To understand, as it were, the darkness of not-being into which the interlocutors throw him and, by the act of throwing, are thrust into themselves.

If we question the psychoanalytic notion of parricide we may come to see that the definitions of the sophist as offered by the interlocutors represent the possibility of satisfied desire. However, this satisfied desire (knowing the sophist) remains closed off to the interlocutors who remain captive within Parmenides' law, using as they do signifiers intended to achieve actual knowledge of the sophist which merely marks a limit for their desire (to know). Conceptualizing the interlocutors' attempts to know the sophist from within the framework of the impossible relation between the Subject and the law, it is made clear that the knowledge they seek cannot, in fact, positively represent the sophist but only represent their desire to know him. The reason for this goes back to what has already been discussed here above—the object of knowledge is, within this structure, also the object of desire which is in fact an object of lack. The Subject's desire (for knowledge) is, alas, the Subject's way of both constructing and regulating his desire which is captured by the taboo.

---

1 It is Socrates that deals with the sophist's self contradiction in Euthydemus. But unlike the Stranger, which criticizes the logic of the argument, Socrates criticizes its ethical effects; the denial of refutation, ignorance and mistake.

2 Heidegger discusses the part in the Sophist wherein the interlocutors violate Parmenides' law and claims that the interlocutors remain, after the transgression, still within the limits of the investigation of Being (Heidegger 1997: 299). Heidegger’s claim is that the subversion of the opposition between being and not-being does not exclude the interlocutors from the limit of Parmenides' law.
Conclusion

I shall end this essay by returning to the opening phrases of the Sophist in hopes of further clarifying the claims made here throughout. The Sophist opens in a rather calm manner, the interlocutors noting the following about the sophist:

'[He] is not the easiest thing in the world to catch and define, and everyone has agreed long ago that if investigations of great matters are to be properly worked out we ought to practice them on small and easier matters' (Sophist 218c; 271, 273).

It would seem 'easier matter' refers to the art of knowing the sophist through acts of deduction and negation; the art of analyzing the 'angler' with which to 'fish' the sophist. Yet it seems that even at this early stage the anticipation of entrapment is already present; the fundamental aporia, of actually succeeding in distinguishing between the philosopher and the sophist is very much present. Embedded in the metaphor of the angler and the fish lie the questions of who is the fish and who is the angler. It is made clear quite early on that Theaetetus is the fish which the Stranger attempts to 'fish' for answers to the questions he raises during the dialogue. Only that the Stranger himself is simultaneously a fish himself, devoting his discourse to escaping the sophist’s metaphoric fishing rod. But perhaps Theaetetus is the angler trying to fish for knowledge of the sophist from the Stranger? And the sophist, the object of this dialogue, is he not the fish that the Stranger and Theaetetus seek to trap? Does this not make both the Stranger and Theaetetus sophists of sorts too—anglers that prey upon their objects only to find themselves trapped in a metaphor of their own making?

Finally, perhaps it is the sophist who is the reigning angler, ultimately trapping both the Stranger and Theaetetus that are after him, for in the process of attempting to capture him they are invariably locked in his realm—hooked in the fishing rod of the one they attempt to hook.

Bibliography

Revisiting the classical definition of art as mimetic practice, I reevaluate Plato’s decree that art, because representational, is removed from truth. Yet I argue, and go on to confirm, that the Theory of Forms, that according to which Plato ostracizes art from truth, is itself an epitome of representation.

Ever since the epistemological consequences of Platonic ontology were systematized in Πολιτεία (V-VII), representation has been cognitively downgraded on the basis of its alienation from truth. It can be demonstrated however that the criterion of truth used to arbitrate in this case is logically reducible to the concept of representation; and it is unjustified to downgrade art because it is representational according to a concept of truth that is itself representational.

With the inauguration of Western metaphysics art is dissociated from truth. Danto refers to this phenomenon as the ‘philosophical disenfranchisement of art’ which signifies a strategy of ‘immunizing’ art from knowledge and receives its epistemological legitimation from the critique of art undertaken by Plato in Republic X. Philosophy, Danto goes on to argue, is co-referential with the disenfranchisement of art. Alarmingly, the strategy of alienating art from truth, thereby rendering it epistemologically valueless, happens simultaneously with the establishment of philosophy as an intellectual discipline.

Significantly, however, it is not art that Socrates indicts in the Republic but rather the process of representation (mimēsis). Yet it is evident from the discussion that Plato considers all artistic practices to be ultimately reducible to representation (his reliance on painting to establish the essence of mimēsis is marked). Thus, despite what some commentators have attempted to argue,

1Danto (1986).
3The phrase ‘alienation of art’ derives from Bernstein (1992).
4Rep. 595a. See also the crucial interrogation of Homer at 599d which explicitly states that the objective of Book X is to provide a definition of representation.
5Moss, for instance, argues that Plato meant to critically condemn only art practices that imitate things as they appear (f-instances and their reflections) and not as they are (F-nesses); this thesis, which Tate and Nehamas share is based on a spurious distinction (with no textual evidence in Plato) between imitation and imitativeness which argues that Plato criticises only imitativeness (defined as enjoyment, anachronistic in the context, in the imitation ‘for its own sake’). Moss (2007) 437. Plato insists that Forms cannot be copied by mimetic means at all because that which imitates the Forms in this manner is the visible world, i.e., the objects of
Plato intends to impugn all artistic practices as epistemically deficient. As this is most obvious in the case of visual art (although not his principal target), this provides Plato with the standard for his assault. As Socrates’s rationale for rejecting art relates to its representational nature, therefore, I will begin by revisiting the Platonic identification of art as mimetic practice (*mimētike technē*) and because mimetic, as thrice removed from truth (*triton ti apo tēs alētheias*).¹

Cross-questioning this uncompromising critique of representation reveals a curious fact, a fact that, amazingly, has remained concealed from generations of philosophers: the criterion of truth against which Plato evaluates representation is itself logically reducible to representation. In other words, the Theory of Forms, the metaphysical apparatus according to which Plato alienates art from cognitive and moral value, is itself an epitome of representation. The philosophical disenfranchisement of art is therefore legitimated on the basis of a conception of truth that is itself implicitly representational.

**The Platonic Critique of Representation**

In the *Republic* 596b, Socrates illustrates the relationship between intelligible Form and its sensible instances using an analogy: the manufacture of equipment. Thus, for example, a ‘maker’ (*tektōna*) of furniture, when manufacturing beds, follows the appropriate Form (“Bedhood”) as a model. The products are obviously not the Form but rather constitute physical duplicates of it that partially correspond to it, products that approximate but never fully epitomize the nature of Bedhood itself.

*If, then, what he makes is not “what a bed really is”, his product is not “what is,” but something which resembles “what is” without being it.*²

Only the Form (*eidos*) “is” in the consummate Platonic sense: it IS unique, unchanging, intelligible and paradigmatic. Each Form, that is, constitutes the consummate archetype (*paradeigma*) of its several empirical instances³ and each constitutes the original (and causal) exemplar of its sensibly perceived physical tokens.

*So we shan’t be surprised if the bed the carpenter makes is a shadowy thing compared to reality?*⁴

---

¹ *Rep*. 597e, 599d, 602c.
³ *Phd*. 74a-74d.
⁴ *Rep*. 597b.
When contrasted with their essential natures (\(\textit{eidē}\)), the objects of physical reality inevitably appear insubstantial, deficient. At this point, under Socrates’s influence, Glaucon learns to distinguish between appearance and reality (i.e., he begins to \(\textit{philosophize}\)); and it is only on the basis of this distinction, key to the dialogue (for it marks the beginning of the dialectical reasoning that ultimately facilitates apprehension of Forms), that the central question of representation can be adequately stated.

Given all that has been said, how do things stand with representation? If the status of representation is to be investigated within the framework of Platonic ontology the question becomes: what kind of reality does the representation (the product of \(\textit{mimēsis}\)) belong to – if the relevant objects that constitute its subject matter are themselves derivatives of original entities, more real (and therefore closer to truth) than them?

In addressing the question, Socrates presents a triad of different manufacturers: the \(\textit{demiourgos}\), the carpenter (\(\textit{tektona}\)), and the painter (\(\textit{zōgraphos}\)); and with them he correlates the subject matter relevant to their practice:

1. The \(\textit{demiourgos}\) fashions one (and only one) bed, the Idea of Bedhood, which, along with unity, possesses all salient predicates associated with the Platonic \(\textit{eidos}\);¹
2. the \(\textit{tektona}\) makes physical, visible versions of this Bed-itself by attuning noetically to its intelligible \(\textit{eidos}\);
3. the \(\textit{zōgraphos}\) constructs facsimiles of the artisan’s physical beds and thus relies on these products and how they appear without relating directly to the \(\textit{eidos}\). That is to say, the painter, unfamiliar with the Idea, imitates visible reality, not as it is in itself but rather as it appears.²

From 1 and 2 it follows that, in contrast to the artisan’s attunement to the \(\textit{eidos}\), the visual artist engages only with the (partially real) empirical token and reproduces only how it appears (or, to be exact, the painter depicts its visible surface aspects).³ Ineluctably, the painter is not acquainted with the true shape and nature (\(\textit{eidos}\)) of the bed itself – nor yet its physical dimensions – but is rather retarded at the level of its shadow-like phenomenal aspects that make it appear different (from various perspectives) and at certain times.⁴

Familiar only with the dubious phantoms of the phenomenal, how can the artist

¹It is intelligible (\(\textit{Phd.}\) 65c, 79a, 80b; \(\textit{Rep.}\) VI, 509), changeless, (\(\textit{Phd.}\) 78d; \(\textit{Rep.}\) VI 484b, 479e), not qualified by contrary predicates (\(\textit{Phd.}\) 74c; \(\textit{Smp.}\) 210a-211a; \(\textit{Rep.}\) 479a-c), paradigmatic (\(\textit{Phd.}\) 74c; \(\textit{Smp.}\) 210a-211a; \(\textit{Rep.}\) 479a-c), and responsible for infusing phenomena with their portion of being or cause of being (Vlastos [1954], 324; \(\textit{Rep.}\) 484c-d; \(\textit{Phd.}\) 76d-e). See also \(\textit{Phd.}\) 100c; \(\textit{Prm.}\) 130b, 131c-132b; Aristotle \textit{Metaphysics} A6 987b-989a and M5 1078b-1080a. In the \textit{Timaeus}, Plato claims that the sensible realm was created by the \(\textit{demiourgos}\) (\textit{Tim.}\) 27d).
²\(\textit{Rep.}\) 597b-c.
³\(\textit{Rep.}\) 598b
⁴\(\textit{Rep.}\) 601c.
have any knowledge of the true nature of being? This is the conclusion Socrates gets Glaucon to draw through the series of elenctic questions from 596a to 598d …

when the painter makes his representation, does he do so by reference to the object as it actually is [to on] or to its superficial appearance [phainomenon, ὁς phainetai; the appearance as it appears]? Is his representation [mimēsis] one of an apparition or of the truth [phntasmos ἐ αλήθειας]? Of an apparition.¹

The situation for the visual artist is however worse than it appears. If the subject matter of the visual artist is confined to shadows and images-of-images (phantasmata) then her work can only ever be regarded as an inadequate representation of what is already an inadequate version of the true nature of reality. Representation, according to the Platonic metaphysical scheme, is always dissociated from truth; but the representation of what is already a representation (a still-life image, as it were, of becoming) is thus even more seriously alienated. Hence Plato’s decree that art practices, because they involve derivative mimēsis of what is already a kind of mimēsis, produce products thrice removed from truth.²

Hence the epistemological fate of the artist: condemned either to represent the visible objects of sensible reality and thus reproduce the dubious simulations of perception (which as Panofsky observes amounts ‘to a pointless duplication of the world of appearances’) or, more charitably, to imaginatively engender ‘unreliable and deceptive illusions’ that ‘mislead our imperfect eyes’ and hence augment moral confusion, the truth-value of representational art is clearly ‘less even than the world of appearances’.³ Visual art, as Plato figuratively puts it, is an ‘inferior child born of inferior parents’.⁴ Art, according to Plato, is a veritable bastard. Worse even than the despised sophist, with the intellectual capacity of the cave-dwelling troglodyte, artists are condemned along with their entire craft to trade in semblances, trivial facades, falsities.⁵ Artists can possess neither knowledge nor belief nor even correct opinion about their subject matter.

Analysis of the Platonic Account

In the first philosophical critique of visual art, therefore, Plato confronts art with the criteria of his ontology and finds it deficient. Besides relying on the assumption that representation (mimēsis) constitutes the essence and purpose of art, his critique also depends, not unexpectedly, on the dualism of the

¹Rep 598b.
²Rep. 602.
³Panofsky (1968) 5.
⁴Rep. 603b.
⁵Rep. 602a.
intelligible and sensible, a dualism, it should be pointed out, ultimately grounded in Plato’s normative construal of the distinction between appearance and reality.¹

**Figure 1. Platonic Ontological Dimensions**

<table>
<thead>
<tr>
<th>DIMENSION ONE</th>
<th>(The Intelligible Realm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line Stratum: L3 / L4</td>
<td></td>
</tr>
<tr>
<td>Epistemic Status</td>
<td>νοησις = διάνοια (L3) / επιστημή (L4)</td>
</tr>
<tr>
<td>Maker</td>
<td>δημιουργος</td>
</tr>
<tr>
<td>Ontological Order</td>
<td>το οντό νοητον</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DIMENSION TWO</th>
<th>(The Visible Realm A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line Stratum: L2</td>
<td></td>
</tr>
<tr>
<td>Epistemic Status</td>
<td>δοξα / πιστις</td>
</tr>
<tr>
<td>Maker</td>
<td>τεκτονα</td>
</tr>
<tr>
<td>Ontological Order</td>
<td>το οντο ορατον</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DIMENSION THREE</th>
<th>(The Visible Realm B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line Stratum: L1</td>
<td></td>
</tr>
<tr>
<td>Epistemic Status</td>
<td>δοξα / εικασια</td>
</tr>
<tr>
<td>Maker</td>
<td>ξυγραφος</td>
</tr>
<tr>
<td>Ontological Order</td>
<td>φαντασμα / σκεια / ειδολα</td>
</tr>
</tbody>
</table>

However, it is important to observe that Platonic ontology is more complex than the two-world dualism standardly attributed to him. According to the topography delineated in *Republic* Books VI, VII and X, there are, apparently, not two but, strictly speaking, three distinct dimensions of existence (Figure 1)

¹Moss (2007) also draws attention to this: the link between the epistemological critique and the moral condemnation hypothesis is explained by Plato’s application of ‘the appearance/reality distinction to matters of value’ 423. See also G. Fine (1999).
(‘there are three sorts of bed’). Beneath the intelligible region, Plato subdivides the sensible region into two further categories having relatively distinct ontological status:

1. The category of physical objects;
2. the category of shadows, images, and reflections of physical objects, including works of visual art.

One sub-section of the visible world, he declares, is made up entirely ‘of images’. Each dimension – from intelligible reality to token instances in the world of the senses, to their corresponding images, shadows and reflections – diminishes in respect of the proportion of being invested in it. As a result of its relative distance (or genealogical removal) from the Forms, we discover that each successive stage in the declension of being, from intelligible through sensible to phenomenal, is bereaved of a crucial percentage of truth.

First there is what truly ‘is’ (alēthōs) or ‘is a true being’. This is intelligible being, the only dimension that counts as real in the emphatic metaphysical sense: it is where what in truth is (ousia esti) ultimately or perfectly real (teleōs on / ontōs on) ‘unaffected by the vicissitudes of change and decay.’ The world of Forms is ultimately and genuinely real as opposed to what appears (to be real) but is not.

Second, material entities and properties, the perceptible denizens of the world accessible to the senses “are” only in a derivative sense: that is, they appear to be real, but, being the diminished semblances of intelligible Forms, are twice removed, and thus already far from consummately real. Plato does not, as pointed out, deny the sensible world all claim to reality; rather he emphasizes that empirical particulars possess a secondary ontological status in relation to their intelligible antecedents. Although participating to some limited extent in true reality, the objects of visible reality are nevertheless ontologically derivative in relation to their paradigmatic progenitors; these, they more or less approximate, but ultimately fall short of. Striving toward true being, material entities only manage to imperfectly embody – in the attenuated, artificial colors of visual simulation – the consummate reality of Forms. According to the metaphysical picture Plato commits to therefore, the empirical world, entirely familiar, and generally accepted as constitutive of reality in the strong sense, is merely a pallid derivative of a more fundamental ontological dimension.

Underneath these strata lies a lower level. Here we descend to the most etiolated dimension of reality where shadows and phantoms, the impressions of the visible world’s tangible objects, converge. Strictly speaking, however, this

---

1 Rep. 597b.
2 Rep. 510a.
4 Rep. 485b.
5 Phd. 74d.
6 Phd. 75a.
zone constitutes a sub-stratum of the visible world (*to horāton*) and therefore there seems no warrant to attribute a *third* ontological dimension to Plato (Figure 1). Moss, for instance, tends toward this error when she maintains that two autonomous ‘distinct objects’ exist, namely, the physical bed “itself,” on the one hand, and the appearance of the bed, on the other. If there yet remains a sense in Plato’s thought in which the phenomenal realm maintains a relative autonomy from the physical tangibility of the empirical world, this is ultimately treated as a mistake of perception that stands in need of *correction*. Correct, *critical* perception,¹ will eventually appreciate that the appearance of the bed is always *its* appearance, that is, the images, shadows etc., of the bed *cannot* be ontologically autonomous from the particular physical bed (even though they *may appear* to be). In fact, as discussed below, the appearance of the bed remains intrinsically coupled with the particular bed as effect is coupled with cause.² This means that although the physical bed may be ontologically distinct from its appearance (because, strictly, they have different properties), they are not *logically* distinct.

Rather Plato is more concerned to demonstrate that, at this stratum, things are rarely discerned as they *are* but rather as they *appear*³ and how they appear not only conceals but actively *distorts* how they *really* are. The ‘phenomenal’ realm possesses the peculiar potency to reverse its effects by weaving a convincing simulacrum of ‘true reality’.Appearances have the power to cast an oneiric spell inculcating widespread illusion that the phantasmagoria of the phenomenal represents an adequate picture of reality.⁴ Thus conveying the seductive message that its recycled imagery is how, and only how, reality, as such, appears: ‘The latter is, as it were’, or, at least appears to be, Heidegger comments, ‘the whole world.’⁵ Having the power not only to deceive but to seduce and stupefy, the phenomenal spectacle works like a cognitive narcotic.⁶ False pictures of reality are created that appear to be faithful imitations – images having the semblance of truth. Hence Socrates correctly exhorts vigilance when it comes to the image: the simulacrum, he teaches, is illusory: it cannot be trusted. What the populace imagines to be real is merely a ‘skiagraphia’, a shadow-play, but yet, at the same time, ‘something’ strangely convincing, ‘akin to trompe l’œil.’⁷ Plato, Moss comments, ‘wants to make the point that realistic [representation] has a certain power over us … the power to *deceive*.⁸

¹See Rep. 595b where Plato introduces the idea of critical perception in terms of an ‘antidote’.
²Rep. 510a.
³Rep. 598a.
⁴Rep. 598c-599.
⁶Rep. 492 b-d (corresponding passages in 599c-601b).
Reality, Representation and the Theory of Forms

If the Theory of Forms involves, in Vlastos’s phrase, a ‘Degrees-of-Reality’ hypothesis (according to which, if the intelligible zone is paradigmatically real) then the sensible world occupies a kind of half-real zone of becoming, ‘between the purely real and the totally unreal.’ On the basis of this, therefore, the lower dimension of appearances, although admittedly close to non-being, ought not to be regarded as totally unreal (Figure 1). Plato, even though he may restrict the full implications of being from the phenomenal dimension, as considered, does not withhold all reality from it.

In this section, following the degrees of reality hypothesis, I argue that the Theory of Forms relies on the concept of representation for its explanatory cogency. As Plato describes it the ontological ‘degrees’ mapped out by the schema of the Intelligible-Sensible-Phenomenal are structured according to the logic of representation. That is, exactly as the images, shadows and reflections of phenomenal reality are, strictly speaking, representations coupled to their more real physical objects, so the visible realm is itself, in its stead, a simulation (or representation) of its true intelligible foundation.

In the Phaedo, Plato unequivocally states that the visible particulars that constitute the physical instantiations of Forms are ‘imperfect copies’ of them. Indeed, the doctrine of recollection is dependent on the logic of representation for its plausibility (as is the related ‘argument from imperfection’). ‘If all these absolute realities, such as Beauty and Goodness’, Socrates says, ‘really exist’, then we re-discover our latent knowledge of them by referring to the items of perceptual reality ‘as copies to their patterns.’ If it is accepted that anamnēsis is caused by the noetic intimation of something through the perception of something else, then sensible objects can be considered, precisely, as signs of their intelligible paradigms; for they act to stimulate the awakening of the latent Form-memories through being representatives of something more fundamental. This implies that the connection between sense-perception (‘aisthesis’) and the memory that facilitates the apprehension of the Forms is a function of representation. Although necessarily exceeding every measure of sense-perception, noetic apprehension is yet in some sense contingent on the correct (critical) perception of sensible reality and the correct critical perception of sensible reality is to perceive it as a mere representation of intelligible reality. I will explain.

Throughout the Republic Plato explicitly identifies the objects of sense-perception as ‘likenesses’ or ‘semblances’ of Forms: thus the sensible particular

---

1Vlastos (1954) 335.
3Phd. 75b my emphasis.
4Phd. 76d.
5Phd. 74c.
is described as ‘something which resembles “what is” [but] without being it.’¹ In the ‘Divided Line’ episode of Republic Book VI (509d-511e),² the graph used to illustrate the epistemological consequences of Plato’s ontology (Figure 2), the visible order is partitioned into a region of simulacra (eidōloi) ‘of physical objects’ and a region where the original ‘physical objects’ (qua original prototypes of these images) reside:³ ‘by “images”’ Socrates says, ‘I mean first shadows [skias], then reflections in water and other close-grained, polished surfaces’. He then asks a question. ‘Would you be prepared to admit that these sections differ in that one is true [alētheiai: genuine], one not, and that the relation of image [to homoioīthen: “the likeness”] to original [to oi homoioiōthē: “that of which it is a likeness”] is the same as that of the realm of opinion to that of knowledge?’⁴ But this alternative perspective is precisely directed at perceiving the items ‘of the visible order in their turn as images’ OF something else.⁵ Pursuing the intelligible causes of phenomenal reality therefore involves adopting an attitude to perceptible objects that regards them, according to the distinction between appearance and reality, as images OF their original Form. If, that is, we adopt a standpoint vis-à-vis sensible reality that is analogous to the faculty that discriminates an original from its representation, Plato suggests, then the qualities of the visible world will disclose their true ontological status as the diminished ‘images’ OF their paradigmatic Forms. Yet this entails treating the objects of the sensible world, precisely, as representations OF the intelligible world. Just as ‘geometers’ employ figures (eidōloi), and thereby rely on designs as graphical representations of ‘the originals which they resemble,’ Plato explains, so too we must rely on images but only in order to think about the originals they resemble (without actually being them). ‘The actual figures they draw or model, which themselves cast their shadow … these they treat as images only, the real objects of their investigation being invisible except to the eye of reason.’⁶

Ultimately, Plato’s Forms relate to the visible entities of empirical reality as physical objects, described at Republic 510a-b as those things that ‘likenesses are likenesses of’, relate to their images. According to the Divided Line, therefore, L1 (images + illusion) is to L2 (perceptual objects + belief) as L2 is to L3 (mathematical Forms + reasoning) as L3 is to L4 (intelligible Forms + noesis); that is, that the relationship of image to physical object (i.e. the gain of a degree of reality) is commensurate with the relationship between physical object and its corresponding Form in the intelligible realm.⁷ Therefore, any

¹Rep. 597a.
²For diagrammatic explication, see the Penguin version of Lee’s 1955 translation, The Republic, 249-251. See also Fine’s (1999) analysis of Republic V-VII, ‘Knowledge and Belief in Republic 5-7,’ especially 229-245.
³See also Fine (1999) 229.
⁴Rep. 510a, my emphasis: ‘as the opinionable is to the knowable [i.e., as the perceptible realm is to the realm of Forms], so the likeness is to the thing it is like.’ This is Moss’s translation (and her parenthetical remark). Moss (2007) 420.
⁵Rep. 510b, my emphasis.
⁶Rep. 510d
physical relationship between a perceptual object and its semblance (shadow, picture, etc.), as exemplified by the visual artist’s mimetic reproduction (picture-bed) of the physical bed, corresponds to the metaphysical relationship between intelligible Form (“bedhood”) and its semblance (the physical bed). The ontological relationships depicted in the line, therefore, can be finessed in terms of physical representation, on the one hand, and metaphysical representation, on the other.

**Figure 2. Divided Line***

*L4*

INTELLIGIBLE
REALM
to νοητόν

*L3*

Forms

*L2*

Physical Objects

*L1*

Σοματικά, Σκιάς, Εικόνα

*Plato recommends that the Line be divided *unevenly* with the greater proportion reserved for the intelligible realm. However for reasons of legibility I have subdivided it evenly.

Of course, the Line’s schematic distinction between image and original is configured vividly in the metaphor of the cave (*Republic* VII 514a-518b). To depict the teleological ascent from illusion – through opinion and reason – towards noesis, given schematic treatment in the Line, Plato constructs the most celebrated thought experiment in the Western tradition. ‘Imagine’ he
begins ‘an underground chamber like a cave, with a long entrance open to the daylight and as wide as the cave.’ Inside, a subterranean audience of inmates, constrained so that turning is impossible, faces the dimly illumined inner wall of the cave, where a shadow-play projected by objects carried in front of a fire in the gloom behind them, is cast. If the captives have been there since birth, Plato conjectures, then they will assume (knowing no better) that these flickering silhouettes on the surface of the wall are all that exists: ‘they would believe that the shadows of the objects … were the whole truth’ (to alēthes). Cast shadows are thus (mis)taken for real things. ‘As long as one encounters nothing but shadows, these hold one’s gaze captive and thus insinuate themselves in place of the things themselves.’ The troglodytes, Heidegger’s exegesis clarifies, ‘do not consider what they have before them … to be a semblance of something else.’ They do not appreciate, that is, the distinction between shadows and the things that cast them. Certainly, they see the shadows, Heidegger remarks, ‘but not as shadows of something.’

If, however, a captive was forced somehow to turn around: what then? Firstly, such a turning would prove a difficult and even ‘violent’ kind of liberation: for ‘he would need to grow accustomed to the light before he could see things in the upper world outside the cave’ he suggests. ‘Those who before looked only at shadows’, Heidegger glosses, will ‘now come mallon ti eggvjerō tou ontos, “a little nearer to what is”’. Some time after the struggle of this difficult rotation has abated, however, the former captive ‘will [now] be most successful in distinguishing shadows: then he will discern the reflections of men and other things in water, and afterwards the realities’ that cause these images and reflections. Finally, we learn to ‘discriminate systematically between the shadows and the objects, and to see that the latter are “more real”’. This ultimately marks the inauguration of the cognitive capacity to distinguish ‘between the … image of an object and the object,’ that is, ‘between appearance and reality.’

The purpose of Plato’s metaphor, of course, is to model what has been delineated more abstractly in the Line. The cave, although itself merely an image, is, Socrates stresses, intrinsically related to his previous teachings. Thus the narrative is intended to picture the mediation between the visible and intelligible worlds, according to the ontological relationships we are now cognizant of. The metaphor is specifically structured, that is, to correspond to the relationship between the sensible world (the ‘realm revealed by sight’) and the intelligible world. The fire in the cave corresponds to the sun that endows the things of visible reality with their availability to sight (i.e., lights up their visible shapes or forms), these, in turn, correspond to the “images” of the Forms: the objects that cast the shadows in the cave are thus likened to the

---

1 Rep. 515c.
5 Rep. 516a.
Forms casting their impressions and thus (in)forming the visible world with its portion of being.

We acquire truth, and reality is revealed, when we are forced to turn around. Only by virtue of this technique is the cause of the flickering chiaroscuro revealed ... dis-covered. What requires emphasis here however is that the turning (periagogē) movement Plato refers to – that is, the orienting procedure that allows us to appreciate shadows as shadows of something by looking behind the scenes, so to speak, at their causes – is the very same cognitive achievement that enables us to distinguish between image (representation) and its original (what is represented). Turning around, that is, away from the sensible world is the very activity, paradoxically, that reveals its foundational ontological basis. Therefore, it is by virtue of this critical act of turning, which, in fact, involves our entire being or ‘essence’, that we learn to abstractly “perceive” the intelligible beyond the perceptible dimension in a manner that corresponds to the turning that allows us to discern originals (reality) behind artificially produced simulacra (appearance).¹

Again, reality in itself, according to Socrates, relates to what appears as original relates to its (mere) representation. Of course, Plato concludes that the philosopher, having successfully turned around, must prepare to re-turn to the underworld. Only now, equipped to appreciate the proper relationship between image and object, she is in a much more influential position to instruct (paidia) others in the art of turning around to discover the shadows’ causes, and, beyond this, to apprehend the ultimate intelligible cause of visible reality itself.

Plato’s emphasis on the turning movement makes sense only if his metaphysics is considered as tacitly governed by the logic of representation; for the turning movement he recommends simultaneously urges us to turn away from the representation to focus on what the representation represents. It is in respect of this cognitive turning, I argue, that the representation is recognized precisely as a representation OF something distinct from that which it represents. Now, as we know, it is logically impossible to confuse a representation with what it represents. Yet this (a priori epistemic intuition) is precisely what Plato relies on in defense of the Theory of Forms. It is so deep and intuitive that no one has ever considered questioning it in precisely the terms that serve to make it explicit. For if it were, then the Theory of Forms would emerge as itself predicated on the very thing it is used to ostracize (for epistemic and moral reasons) in the Politeia.

Truth and Representation

What is the criterion of truth such that Plato can claim with confidence that representation is removed from it? In this context truth is synonymous with

¹This interpretation indeed simplifies matters, for it omits to consider that the final epistemic graduation into noesis proper is totally independent of the images of the visible dimension and pursues its inquiry solely through apprehending the Forms themselves (510b). Reaching this stage however, is totally contingent on the earlier stages.
‘true being,’ i.e., the Form considered independently of its physical instances. The criterion of truth is thus the ontological causal principle of visible phenomena – as cognitively apprehended.

Truth for Plato, to put it crudely (but accurately), is identified with proximity to the Forms. But what is the modality of this proximity? Proximity to the Form must either mean a cognitive closeness, as sketched in the Divided Line, or, if Plato would insist on a metaphysical sense of nearness, then this proximity must have direct epistemological consequences. Thus the criterion of truth becomes epistemic proximity to the Forms: ‘there are, corresponding to the four sections of the line … four states of mind’ Socrates says, ‘you may arrange them in a scale, and assume that they have degrees of clarity corresponding to the degree of truth possessed by their subject-matter.’

When its metaphorical resonance is removed, the criterion of truth as nearness emerges as a kind of conformity and, therefore, ultimately, a kind of correspondence; indeed, the Platonic locution, το ὑπό τοῦ, very strongly implies a normative structural affinity between mind (nous) and Form.

And of course it is here that the concept of truth first begins to assume the contours of adequacy (homoiōsis/adaequatio) that proceeds to determine the treatment of truth in philosophy. Indeed, according to Heidegger, it is precisely in Plato’s Republic that the ‘essence of truth’ (understood in the original Greek alpha-privative: a-ληθεία) which has, incidentally, he claims, a fundamental relation to the image of a subterranean cavern, undergoes the crucial transformation from ‘unconcealment’ to homoiōsis (likeness) or, as he prefers to parse it, “agreement.” And agreement, Heidegger acknowledges, here means epistemic representation of the object of knowledge.

Turned toward being and away from the shadows, the captive in the cave, as considered, sees ‘more correctly’. ‘In so directing itself’ Heidegger comments ‘apprehending conforms itself to what is to be seen: the “visible form” [eidos / idea] of the being. What results from this conforming of [or orientation towards] apprehension’ he concludes ‘is a homoiōsis, an agreement of the act of knowing with the thing itself.’ This transformation in the conception of truth from unconcealing αληθεία to homoiōsis (correctness of representation) constitutes a benchmark episode in the history of Western philosophy.

Our analysis of the problem of representation in Plato has therefore discovered that the Theory of Forms, so hostile to the activity of mimesis, is itself paradigmatic of the object of its hostility. The conclusion is inevitable: the Platonic critique of representation which condemns visual art as alienated from truth is ultimately established on an implicit theory of truth as representation (homoiōsis, adaequatio). In Greek, mimēsis signifies “representation,” i.e., the relationship of likeness between what is represented

---

1Rep. 511e.
2See Panofsky (1968), where he identifies what he calls ‘cognitive truth’ with ‘correspondence to the Ideas’ (4).
3Rep. 515d.
and its representation. Yet as considered, Plato uses the noun *homoïôsis* to denote the concept “likeness” pertinent to the relationship between his ontological categories. Andrew Benjamin, through a reading of the account of nomenclature in the *Cratylus*, has drawn attention to the conceptual affinity of *mimēsis* and *homoïôsis* in Plato’s thought: ‘the name (to onoma)’ he says ‘is an imitation (mimēma) of the thing named’. Here, the hypothesis that *mimēsis* is a general form of a more specific structural correspondence or ‘likeness’ (*homoïôsis*) is thus presupposed.¹ In both cases, the relationship is one of *homoïôsis* which is conceptually affiliated with *mimēsis*. Is it possible, therefore, to translate both terms, salva veritate, as “representation”? Benjamin suggests so.

The classical definition of truth as the correspondence of ‘intellect and thing’ (*adaequatio intellectus et rei*) constitutes the infrastructure of what Heidegger terms the ‘normative conception of truth.’² This in turn can be reduced to the “correctness of representations,” a conception that determines *a priori* all philosophical treatment of the theme of truth. From Peripatetic metaphysics to post-Cartesian epistemology,³ from logical positivist truth-functionality to Tarski-type disquotational equivalence schemas, truth is invariably considered in terms of adequacy; a species of correspondence logically reducible to the *correctness of representation*. Yet if, as Heidegger argues, truth is thematized on the basis of a prior, assumed to be intuitively obvious, concept of representation, then to judge art as epistemically compromised according to a criterion of truth that is itself, just like art, representational is arbitrary and unjustified. A project of transvaluation is hereby suggested that aims to retrieve art from philosophical disenfranchisement by resurrecting the fundamental form of truth-disclosure as phenomenal emergence from its genealogical dormancy. Truth in this sense is identified as ‘a bringing forth … out of concealment and specifically *into* the unconcealment of its appearance.’⁴ And, as Plato was well aware, art provides the conditions of possibility for this event of *alētheia*.

References

⁴Heidegger, *Basic Writings*, 184.


CHAPTER FOUR

Euripides and Plato on Autochthony

Jaroslav Danes

The myth of autochthony was one of the key elements of ancient Athenian democratic ideology or political imagery, which is obvious with respect to the copious references penetrating all classical literature, no matter whether we take historiography, philosophical prose, oratory, drama etc.¹ According to the myth of autochthony, the citizens of Athens, or rather the ancestors of the citizens, were not born of human mothers and fathers but fashioned in the earth. There were several autochthonous beings in Athenian mythical history: Kekrops, the first king of Athens, Erichthonios, the second king of Athens and the first Athenian, and Erechtheus. Erichthonios or Erechtheus are those who, in the language of poetry, from Homer to Sophocles, give Athens its name.² But what are the political functions or implications of the autochthony myth? In principle there were two basic functions of the myth. The first of them pertains to the ideology of citizenship. The myth of autochthony says who is and who is not a citizen of Athens and that citizens are equal because all Athenians can trace back their origins to the first ancestors born of Attic soil.³ Second, it used to be employed as an ideological cornerstone of Athenian foreign policy because it establishes the hegemony or the supremacy of the Athenians over the other Greeks. Only the Athenians, among all the Greek tribes, are the original inhabitants; the other ones are immigrants not born of Greek soil. These two functions are obvious in classical literature.⁴

Therefore, it is not surprising that two clever critics and observers of Athenian society – Euripides and Plato – paid particular attention to this myth. But they paid attention to different aspects of this myth. In this article I would like to answer the question of how they worked with the myth of autochthony.

¹For example, Herodotus 7.161; Aristophanes, Wasps 1075-1080; Plato, Menexenus 245d. For detailed analyses of Athenian autochthony myths see Loraux (1993:3-71). On the political function of myths in general as a symbolic representation of collective unity, see Loraux (1993:37) and Nilsson (1951: 49-80).

²In Homer we hear of Erechtheus’ land (Il. 545-6). Basically, the people of Erechtheus and the people of Athens are synonymous.

³Moreover, there is a further meaning linked to the second point. As Walsch (1978: 301) says: "For an Athenian, to be earthborn (γηγενής) and indigenous (αὐτόχθων) was also to be well-born (εὐγενής), and so his national pride was tied closely to his sense of inherited personal status: if his city was free of racial impurity, so was he." (see Plato, Menexenus 237b) This collective εὐγένεια was used to suspend the class conflict and gave dignity not only to a small group of aristocrats but to all the citizens of Athens. On the establishing of equality, see Plato Menexenus 238d-239a.

Euripides on Autochthony

Euripides made the autochthony myth a central issue of his play Ion. But the story the god Hermes tells in the prologue of this play is a version that is otherwise unknown and Euripides is the only source for it. The plot of the play is complicated and is very similar to motif work of the later New Attic Comedy and novel prose writing of the Hellenistic era. Creusa, the daughter of Erechtheus, has been raped by Apollo. She becomes pregnant but hides her pregnancy, as well as the birth of the child. She leaves the baby in a basket with two golden snakes (the sign of Erechtheus’ house) at the place where she was violated by Apollo. Hermes takes the infant on behalf of Apollo to Delphi, where he is raised by Apollo’s priestess and serves as a temple boy. Meanwhile Creusa is given to Xuthus, a Euboean general, who helped the Athenians during the cruel war with the Chalkedonians and secured the victory for Athens. However, this marriage remains childless, which prompts Creusa and Xuthus to go to Delphi for an oracle. Creusa also hopes that Apollo could reveal the fate of their baby. Apollo surprises Xuthus by telling him that the first person he meets after leaving the temple will be his son. Xuthus meets the temple boy and he calls him Ion. After this lucky and unexpected meeting Xuthus thinks of Ion as a future king of Athens. Creusa is jealous of Xuthus’ new son and plans to murder Ion, being advised by an old tutor who presents himself as an expert on the purity of the Athenian race and the line of autochthonous descent. Fortunately, the temple birds, which are expelled from the temple by Ion, prevent the disastrous plan from being realised and Ion is saved. But Creusa is convicted of committing that crime and Xuthus wants to kill her. Only then does a recognition scene take place. The complications are straightened out by a deus ex machina – the goddess Athena, who announces the glorious future deeds of Ion and his children and sends Creusa and Ion back to Athens to rule it.

We may say that although the action takes place in Delphi it is a very Athenian tragedy. The myth of autochthony and its implications are substantial in every character in the play, from the beginning until the end of the play. Let us examine the play from the point of view of birth and citizenship. Creusa comes from a pure city (καθαρὰ πόλις – 673,1333) ruled by the ancient race of the Erechtheids (469). Even her appearance betrays her noble origins. She is the last legitimate member of Erechtheus’ autochthonous lineage. On the other hand, her pride as regards her origins limits her relationship with her husband Xuthus, who is referred to as a foreigner (290) and warrior (296-8) even by Creusa herself. But Ion’s situation is much worse in this respect because he is a slave (309) and he is supposed to be of base origin (837-8). After having been

---

1I mean the fact that the story depicts Apollo as Ion’s father. On the possible sources for Euripides, see Owen (1939:XII-XIII) and Walsch (1978:311). On the revisions and general reinterpretations of the history of Athens in the 5th and 4th centuries by Athenians themselves, see Pearson (1942 and 1972) and von Fritz (1967).

2Nicole Loraux (1993:184) says: "Athens is the sole subject of Euripides’ Ion, the Acropolis its sole hero. "
given to Xuthus and having finished his investigation of him he seems to be satisfied that he is not base-born. In a conversation with Ion Xuthus betrays himself as a blasphemous foreigner because he denies the possibility of autochthony, which is, on the contrary, an indispensable part of the Athenian discourse of citizenship. He also surprises the very smart and cautious Ion with a new proposal. Xuthus offers Ion the chance to go back to Athens and promises he will secure the position of the king for him. Ion’s reply emphasises the connection between birth and citizenship. He wishes his mother were an Athenian; otherwise he would not be guaranteed the freedom of speech (παρρησία) and his mouth will remain servile (668-675). In short, his father is a foreigner and his mother unknown. According to the law of citizenship, he will be a foreigner in the city even if he is a grandchild of the god Zeus after his father. Exclusion from political society can be avoided only by the authority of his new-found parents but the privileged position promised to him by Xuthus will probably provoke the jealousy and envy of the autochthonous citizens, which Ion rightfully fears. Here we come to a link to the exclusivity of the citizenship. This exclusive society of autochthonous citizens is characterised in Ion’s speech in the following words: fear and envy. In thirty lines these characteristics appear repeatedly: fear (601, 624, 628,), envy (597,611). While fear and envy were the words used for the description of tyranny (or monarchy) from Herodotus (3.82) onwards, here they characterise the egalitarian autochthonous society. According to Ion, it is better not to long for power because the political environment is a sphere of fear and envy. The fears of the autochthonous society are diverse but one of the most distinctive is xenophobia. The purity of the Athenian race, as well as the city, is emphasised throughout the play. Surprisingly, the most passionate advocate of this purity is the old tutor—a servant of Creusa and the former pedagogue of Erechtheus. Surprisingly, because he himself is a slave. He illustrates perfectly why one should fear such a society. He is an embodiment of evil and a master of conspiracy. First, he manipulates Creusa against Xuthus and Ion in the name of racial purity (or for patriotic reasons, if we want). He tries to persuade her that Xuthus plans to thrust them forth from Athens (810-11). In his opinion Xuthus had a baby with a slave woman and sent it to Delphi to be brought up. He had been informed that the child grew up and then he mounted the present trip to Delphi so that he might pretend to find Ion by accident (813-831). Of course, he is supported by the chorus of Creusa’s servants. However, his violent aims are, for a moment, shifted to a new target after Creusa’s unexpected confession of involuntary intercourse with Apollo. He proposes burning Apollo’s temple in

1It seems that Ion alludes to the Athenian law of citizenship. According to Pericles’ decree of 451, the rights of citizenship were conditioned by having both parents Athenians (Aristotle, Athenian Constitution 26).
2Euripides catches a paradox of Athenian democratic society. To stand aside and lead a quiet life (ἀπραγμοσύνη) means not to have any influence at all and, with respect to the normative standards of Athenian politics articulated by Pericles, it means to be useless (Thucydides 2.37-42). On the contrary, to take part in politics implies immersing oneself in the waters of fear and envy.
revenge (974). In principle, he wants to extinguish or eliminate all Creusa’s alleged enemies or those who have wronged her: Xuthus, Apollo and Ion. Creusa initially does not seem to be actively involved in this crazy defence of Erechtheus’ house because she is afraid of never-ending suffering and she also respects the marriage bed (975, 977). She is rather a machine of mourning voices. But, finally, her bitterness pours itself out on Ion, whose murder is plotted with typical cunning, and only then does Creusa present herself as a defender of purity (1036-38) and – cynically speaking – an economy of violence. Yet again the most decisive plea for purity does not come from Creusa but is articulated by the chorus (1058-1060) – it must never be that some foreigner should rule the city instead of the Erechtheids. If the lethal plan fails and Ion survives, the chorus sees only one solution – to commit suicide, because it would be unbearable to watch a foreigner to rule Erechtheus’ city (1069-72). Look what consequences the exclusivity of citizenship combined with the concept of the purity of the race leads to!

Is Euripides’ Ion a pure criticism of the Athenian ideology of citizenship? Did he show its limits? What could the reaction of the audience to Ion be? The shared ideas of citizenship and purity in 5th-century Athens were those which were propagated by the Old Tutor and Creusa in the play. However, it is hardly imaginable that any Athenian could identify himself with those characters – an aggressive and envious slave and a disillusioned and misinformed queen. We may guess that Euripides prompts the Athenians to reject the doctrine of citizenship and racial purity by creating the characters of Creusa and the Old Tutor who they could not identify themselves with and showing Ion – the true Athenian and the son of the last autochthonous being – as a victim of the lethal plans generated by this doctrine. We witness a political thought experiment through a reverse perspective. This experiment might probably reduce the xenophobic inclinations of the audience. The concept of purity is nicely problematised in the case of the relationship between Ion and the birds.\(^1\) At the beginning of the play Ion is angry with the birds for polluting the sacred place (107-8) and trying to build a nest in the temple of Apollo (172-3). He threatens to shoot them with a bow. Paradoxically, these birds will later save his life and lose theirs by drinking the deadly poisonous libation of wine which was prepared for Ion by Creusa and the old tutor. This metaphor shows that the maintenance of purity, regardless of the cost, is not only disadvantageous but could also ruin what it intends to protect. However, the concept of purity, as Walsch says, is constantly challenged throughout the play and Ion’s fate is the best example.\(^2\) Ion’s life is the result of the crime and evil committed on Creusa. Creusa’s pure blood has been mixed with Apollo’s. But can the pure blood be contaminated by the blood of the god? The question of purity returns at the end of the play, when Creusa and Ion exchange their roles. Creusa is now the haunted one and Ion decides to murder her at the altar. Ion suddenly switches to the position of the old tutor and defines purity as killing enemies

\(^1\)See also Walsch (1978:306) and Saxonhouse (1990:273).
\(^2\)Walsch (1978: 306)
Fortunately, he is prevented from doing so by Pythia and thus avoids the double pollution of the blasphemous violation of divine law and matricide.

The last subversive point hides itself at the end of the play. In her speech the goddess Athena changes the traditional genealogies. Ion was known as Xuthus’ son. In Euripides’ fashion he is the son of Apollo. Two versions of the myth have it in common that Ion is the father of the Ionian race, which was a shared belief and not only an Athenian one. But Ion is not a single eponymous hero in Athena’s speech. Creusa and Xuthus will have children, Dorus and Achaeus, as eponymous ancestors of other powerful Greek tribes (1589-94). In the introduction to his edition of Ion Owen suggests that the political message of the play lies in making Ionians proud of their kinship with Athens and showing the Dorians their racial compatibility with the Athenians. As for the Ionians, their kinship with the Athenians was something banal and Euripides’ innovation with Apollo fine but changing nothing. The Dorians might even feel offended that the father of their eponymous hero had been ridiculed throughout the play and that their origin depends on the Athenian Creusa. The patriotic interpretation of the play, stating the supremacy of Athens and showing Ion as ideological support for imperial policy, misses the subversive and critical points of the play. Therefore the political message must lie in something different. Ion was produced during the Peloponnesian war and probably after the catastrophic Athenian expedition to Sicily in 415. What is the action and the result of the play? After the wild turmoil and switching of perspectives, harmony is restored. In the world which Athena sketches in her final speech there is no place for controversies. No controversies can be expected among brother Dorians and Ionians and even the Ionians themselves, because the offspring of Ion’s children will strengthen Athens by gaining control of the islands and the surrounding land (1582-85). It will be an orderly world with a clear distribution of power and also a world of peaceful mutuality and the reciprocity of all the powerful Greek tribes. This vision was contrary to the Greek world at that time: the Ionian allies revolted against Athens, while the Athenians waged a disastrous war with Sparta and oppressed their own allies. The mythically constructed mutuality of the Greek tribes overcomes the fear and envy stemming from the autochthony myth.

---

1Euripides (2002:236) Καθαρὸς ἅπας τοι πολεμίους ὡς ἂν κτάνῃ.
2My opinions are contrary in this point to those of Barron (1964), Gregoire (1923) and Delebeque (1951), who interpret Ion as a classical patriotic play.
3See Herodotus 7.94, 8.44.
4For example, Herodotus (1.147) says that pure and true Ionians are those who are of Athenian descent. Later in his History (5.97) he mentions that the Milesians were settlers from Athens, and that therefore the Athenians should help them in their revolt against the Persian king.
5Owen (1939: XXXIX-XL). There is a very strong and contemporary point against Owen’s suggestion contained in Thucydides’ book. Thucydides argues repeatedly in his History of the Peloponnesian War (4.6, 6.85, 7.57) that neither the Athenians nor their enemies cared for the ties of kinship but acted rather on the basis of expediency and rebus sextantibus.
Plato on Autochthony

Plato handles the autochthony myth extensively in three dialogues, Menexenus, the Statesman and the Republic, but with different degrees of seriousness and impact. For the sake of brevity I will mostly concentrate on the Menexenus dialogue and the Republic.

The Menexenus dialogue is a “seriously meant parody”¹ of Pericles’ funeral oration. At the request of Menexenus Socrates declaims a funeral oration composed by Pericles’ mistress, Aspasia. Aspasia is also said to have composed Pericles’ famous funeral oration (236b). The funeral speech and ceremony were one of the occasions on which the power and beauty of Athens were celebrated². No surprise, then, that the autochthony myth is mentioned in this context because it highlights the Athenian exclusivity. But this, one of the key institutions of the Athenian state, is ridiculed in Menexenus and this parody is a part of Plato’s criticism of political rhetoric (235a-c). First, it would be unimaginable that a woman could deliver such a speech because the public space was closed to women, and second, Socrates’ reference to his teachers, Aspasia and Konnos, who are better to him than Lampros and Antiphon (236a), signifies that he is not serious, because Lampros was a notoriously famous teacher of music, while Antiphon was counted among the ten prominent Attic orators. Despite these ironic accents Aspasia’s speech, declaimed by Socrates, is a standard one and contains all the substantial ideological points, beginning with the autochthonous origin of the Athenians (237c-239a), arguing that Athens is the leader of Greece and the suppliant’s protector³, accounting selectively for Greek and Athenian history, and ending with a celebration of the state that cares for the upbringing of the orphans whose fathers died in battle for Athens. Here autochthony functions in the classical manner. It establishes a collective nobility of birth because the Athenian forefathers were not immigrants but indigenous beings who sprang from the soil and lived in their

¹I use this paradoxical term in order to express that despite a tendency to make something ridiculous, some serious statements and arguments are also expressed. A sign of parody can be seen in Socrates’ evaluation of the effects of such speeches. Fascination and exaltation are the most discernable effects of the funeral speeches. If I listen to them, Socrates says, it seems to me that I am more noble and better than I am and I grow in stature in the eyes of the strangers who listen along with me. This majestic feeling, he says, occupies my mind for more than three days and I hardly remember that I am on the earth even on the fifth day after hearing this speech (235a-c). On the one hand, the phrases and sentences would provoke a comic effect, while on the other one should realise the danger of ideology which covers the contours of the real world.

²On the funeral oration as a political genre and the institution of a democratic city see Loraux (2006).

³This topic is very frequent in Athenian tragedy and oratory. Eight tragic pieces celebrate Athens as a refuge of the suppliants: Aeschylus’ Eumenides, Sophocles’ Oedipus at Colonus, Euripides’ Suppliant Women, Children of Heracles and Medea among others. Cf. Lysias, Funeral oration 7-11; Isocrates, Panathenaikos 168-174, Panegyrikos 54-58, Demosthenes 60.8.
own true fatherland.¹ This nobility of birth involves magnificence of character (245c). The purity of the Athenian race makes Athenians hate barbarians, in contrast to other Greeks who are nominally Greeks but in fact a barbarian blend (245d). The autochthony myth supports an egalitarian society, since all are equal with respect to their origin (ἡ ἐξ ἴσου γένεσις)². Up to this point there is nothing unusual. But what is really surprising is that Aspasia celebrates the earth as a true mother and describes the pregnancy and birth of a human mother as an imitation of Mother Earth (237d-238b). The autochthony myth bluntly neglects woman and pretends that women do not exist. Aspasia thus breaks the rules of mythography.

Here we come to a comparison with the function of the autochthony myth in the Statesman. The character of the Eleatic Stranger in the Statesman tells a story about the age of Cronos and Zeus and the movements of the universe in opposite directions. In the age of Cronos all beings were born of the earth (271a); there were no families, no husbands and wives, but also no cities or states. All these things and institutions originated during the inferior age of Zeus and heterosexual reproduction is one of them (272a-b). In other words, heterosexual relationships and reproduction are part of an inferior world, which is generally a world of boundaries: family, city and state. In this fashion the autochthony myth appears to lose its political implications and there is also no trace of Athenian exclusivity because all beings, including man, come upon the light of the world autochthonously. So, there are no differences in their nobility and origin; nothing separates them from each other. Here the autochthony myth functions in an opposite manner. It does not make some groups of peoples exclusive, but it stirs up boundaries between peoples and even between human and non-human creatures. In general we may say that the autochthony myth in the Statesman occurs in a pre-political world and the unity which it implies eliminates the typical characteristics of the political world: conflict, boundaries etc.³ The autochthony myth in the Statesman, then, stands in opposition to that in Menexenus.

Let us get back to a comparison between the Menexenus dialogue and Pericles’ Funeral oration. Pericles and Aspasia proceed in a different way in order to establish a unity of political society. Though Pericles starts his speech with reference to the forefathers of the Athenians and “an unbroken line of successive generations”⁴ he quickly switches to a celebration of the present. Autochthony does not seem to have any prominent place in his account of Athens. He rather emphasises the value of the Athenian constitution and institutions and democratic ideals of equality and freedom and pictures Athens as a superpower. But who were the dead citizens to whom this oration would have been dedicated, although they were hardly mentioned and what did that

---

¹ Plato, Menexenus 237b: Τῆς δ’ εὐγενείας πρῶτον ὑπήρξε τούδε ἢ τῶν προγόνων γένεσις οὐκ ἔπηλυς οὖσα, οὐδὲ τῶς ἐκείνους τούτους ἀπορησμένη μετακοῦντας ἐν τῇ χώρᾳ ξύλοι θύμων, ἀλλ’ αὐτόχθονας καὶ τῷ ἄνι πατρίδι οἰκούντας καὶ ζῶντας.
² 238e
³ Cf. Saxonhouse (1992:128-130)
speech actually do? According to Saxonhouse, “it is a speech that virtually
denies the death of the dead by making them one with a vital and vibrant city
that lives. The present, an immortal present where past and future are one,
where death is meaningless because the soldiers killed in battle live on in the
memory engraved in the hearts of the citizens.”\(^1\) Pericles blurred the boundaries
between the dead and the living; he abstracted the traits of the real world. He
“invented”\(^2\) an Athens which unifies her citizens in the present and timeless
power and beauty. On contrary, Aspasia’s speech in \textit{Menexenus} is turned
towards history. She starts with the autochthonous origins of the Athenians and
then offers an overview of Athenian history until the present days. Only then
does she move on to speaking of the present conditions. Aspasia makes
autochthony a true source of political unity. This unity is based on the equality
of birth (\(\eta\ \epsilon\xi\ \iota\sigma\omicron\ \gamma\epsilon\nu\epsilon\sigma\omicron\iota\))\(^3\) The equality of birth results in the proper type of
constitution.

“We and our people...being all born of one mother, claim to be
neither the slaves of one another nor masters; rather our natural
birth-equality drives us to lawfully seek legal equality and to yield
to one another in no respect save in reputation for virtue and
understanding (239a).”\(^4\)

We can immediately notice the difference between Pericles and Aspasia.
For Pericles equality is implied by the constitution (Thuc. 2.37.1), which is a
construction, while for Aspasia it comes from the autochthonous and equal
birth, which is something natural. Aspasia describes a constitution as “a thing
that nurtures men” (\(\tau\omicron\sigma\omicron\omicron\iota\theta\omicron\omicron\chi\omicron\omicron\iota\omicron\omicron\omega\omicron\nu\), 238c). If it is good, the citizens are
also good and the city flourishes. If it is bad, the opposite happens. Aspasia
then demonstrates that the constitution of the Athenians’ ancestors was noble
and therefore they were all noble men. Our constitution is pretty much the
same, says Aspasia, and we also share in this nobility. Then Aspasia stirs up all
the distinctions between regimes when she says that some people call this
constitution a democracy and others something else, according to their fancy,
but truly it is an aristocracy (238d). Why is it an aristocracy? That is because
all the citizens are members of the nobility because of their autochthonous
origin, which is noble. We witness an appropriation of aristocratic language to
the equality of (autochthonous) birth. What is, however, strange is the fact that
despite a wide historical account of the Athenian history Aspasia does not draw
a distinction between the periods when Athens was a monarchy or an oligarchy
or a democracy. Thus she says that the type of regime is something secondary

\(^1\)Saxonhouse (1992:120).
\(^2\)I borrow this term from Loraux (2006).
\(^3\)\textit{Menexenus} 238e.
\(^4\)Transl. by R. G. Bury. Plato (2005:347): αλλ᾽ \(\eta\ \iota\sigma\sigmaο\nuι\ νι\μα\ς \ ή\ \κα\τα\ \phi\u0393\iota\ νι\σο\nuι\nι\ \alpha\nu\αγκα\κε\iota\ \zeta\iota\tau\iota\ \κα\τα\ \ν\omicron\omicron\nu\,\ και\ \μη\dnu\ \αλ\llo\ \υ\pi\e\iota\ke\iota\ \αλ\llo\lho\iota\ \ι\ άρε\t\iota\ η\ \d\o\z\iota\ Και\ \phi\ro\nu\iota\sigma\tau\iota\ως.
Euripides and Plato on Autochthony

to true and more deeply-based political unity, which is a bitter irony because the Athenian aristocrats, who were called oligarchs by Aristotle, had always denied this type of unity. So, despite the account of Athenian history this political unity based on autochthonous birth is timeless. But it is a different timelessness from that of Pericles. Aspasia and her speech are, thus, a picture of many paradoxes or even contradictions: a historical account which results in timelessness, a constitution which nurtures men who need not to be nurtured because their autochthonous birth is noble etc.

The use or the ultimate implication of autochthony is completely different in the Republic, where Socrates, Adeimantus and Glaucon think about the establishment of a just city and education in the just city. The exclusive class of guardians who are determined to rule the city must undergo special training and education. They choose the traditional way of education: gymnastic training for the body and musical education for the soul. A musical education is a kind of speech and speech is either true or false (376e). The process of education starts with children, to whom we tell fairy tales about gods and heroes. Their souls are very sensitive and must be cared for carefully. Therefore it is not apt to picture the gods as humble beings who deceive or commit crimes and adultery. The gods are perfect and good and therefore they never lie or commit base deeds. Human beings, on the other hand, are imperfect; however, this imperfection does not imply something bad. Sometimes even a lie is useful and can prevent us from brute evils. But who, when and under what conditions has the right to lie? After painful debate Socrates comes to the conclusion that only a guardian, who is compared to a physician, is allowed to lie in the interest of the state against its enemies or in favour of those who are ruled (389b). Now we come to the autochthony myth; indeed, we must come to it because we are discussing the foundations of political society and foundations always hide some points which are not a matter of fact. How can the unity of political society or the city be maintained? That is the question. Socrates reminds Glaucon of an old story, which he calls the noble lie (γενναῖόν τι). There is a well-known Theban autochthony myth (414c-e) about warriors fashioned in the earth with their arms. It is strange that Plato the Athenian makes use of a Theban myth instead of an Athenian tale. But the Athenian story has egalitarian implications, which Plato himself acknowledges when Socrates says that all citizens are brothers. The Theban myth, on the other hand, has aristocratic implications, which are more attractive to Athenian aristocrats. Maybe this is the reason for Plato’s choice. But Socrates proposes an innovation which is neither egalitarian nor aristocratic in the proper sense. The god mingled the gold in the generation of those who are fitted to rule so that their dignity is greater than that of the two remaining classes, helpers and craftsmen or farmers. The god endowed their generations with silver and iron and copper respectively (415a). After this use of a noble lie Socrates and Glaucon resolve the problem of whether the guardians and their subjects will

1 414c
2 415a ἐστὲ οὐκ ἀδελφοί
tend to believe this story (415d). They are both sceptical but their optimism increases with respect to the following generations. The present generation knows very well the difference between the truth and a lie. But by repetition for many generations this myth should become the truth. This false truth will have a positive effect in making citizens more likely to care for the state and one another. The implication of this myth is a hierarchical society but not an aristocratic society in the proper sense because there is nothing like a hereditary privileged position. For it is not impossible that silver or copper beings would have golden offspring (415b). Anyway, Socrates’ innovation is directed against the egalitarian society which results from the Athenian autochthony myth. Plato, thus, betrays the fact that despite his critical attitude to the Athenian democratic ideological framework, he realises and acknowledges the strength of this instrument.

**Conclusion**

Both Euripides and Plato approached the autochthonous myth and its ideological implications critically. In *Ion* Euripides challenged the picture of Athens as an open and tolerant society by showing the fear and envy which result from the exclusivity of race and citizenship based on the autochthony myth. This limitedness and exclusivity is also an obstacle for the Greek mutuality, which is necessary for orderly and peaceful cooperation. The question of exclusivity is also touched on in Plato’s use of autochthony in the *Statesman*, according to which all beings are equally born from the earth in the age of Cronus. Here every potential exclusivity disappears. Moreover, autochthony exists only in the prepolitical period. On the other hand, in *Menexenus* and the *Republic* we find ourselves in the realm of the political. The key issue is the question of the unity of the political society. Despite the parodical character of *Menexenus* Plato discerns egalitarian implications of the autochthony myth, which he tries to avoid in the *Republic*. Never and nowhere does he question the importance of this myth for the unity of the political society and the question of exclusivity appears not to bother him in the same way as it does Euripides. But as an enemy of the egalitarian society he reshapes the autochthony myth (his famous three classes of citizens: gold, silver, and iron-copper) in order to legitimise a hierarchical political society. In other words the autochthony myth creates an exclusive society but the key question is an exclusivity of what and who, while Euripides seems to warn us against the concept of exclusivity and the limits it produces.

**Bibliography**

Euripides and Plato on Autochthony

CHAPTER FIVE

The Myth of Gyges and the Possibility of Altruism

Chrysoula Gitsoulis

In the opening sections of the Republic, Book II, the character Glaucon sets out to establish the following claims:

(a) Human beings created justice through a social contract, not because they value it for its own sake, but only for its consequences.
(b) (Descriptive claim) When human beings practice justice, they do so against their will, i.e., human beings are by nature unjust. No one is ever just willingly, but only because they are constrained by law (the contract) to be. If self-indulgence could be practiced without fear of punishment, then the tendency for being unjust would prevail.
(c) (Normative claim) The life of an unjust person is much better than the life of a just one, so long as his injustice goes undetected; hence, one is better off being unjust if he can get away with it.

This paper will be devoted to critiquing the descriptive claim (b) about human nature. I will try to show that Glaucon’s thought experiment (The Myth of Gyges), and other popular arguments that have been offered more recently in the literature, do not establish that (b) is true.

Glaucon’s story involving the shepherd Gyges runs as follows. One day, there was a violent thunderstorm as Gyges was feeding his flock, and an earthquake broke open the ground and created a chasm at the place where he was tending his sheep. Marveling at the sight, he descended into the chasm. There he beheld many wonders, among them a hollow bronze horse with window-like openings in it. Peeping in, he saw a corpse, which seemed to be more than human size, wearing nothing but a gold ring on its finger. He removed the ring and made his way up and out of the chasm. He wore the ring at the usual monthly meeting that reported to the king the state of the flocks. And as he was sitting among the others, he happened to turn the stone of the ring on his finger toward the palm of his hand. When he did this, he became invisible to those sitting near him. He was amazed to hear those who sat near him speak of him as if he were absent. Fingering the ring again, he turned the stone outward, away from the palm, and again became visible. He began to experiment, turning the ring this way and that to see whether it indeed had this power – and it did. When he realized this, he at once arranged to become one
of the messengers sent to report to the king. On arrival, he seduced the queen, and with her help, murdered the king. Thus it was that he became king of Lydia.

Now let’s suppose, Glaucon says, that there were two such rings, one worn by a ‘just’ and another by an ‘unjust’ person. No one, he says, would be so incorruptible that he would stay on the path of justice and keep his hands off what does not belong to him, when he could take whatever he wanted from the marketplace with impunity, go into people’s houses and have sex with anyone he wished, slay – or release from bondage – whom he would, behaving toward other men in these and all other things ‘as if he were the equal of a god’. The reason for this, he maintains, is ‘the desire to outdo others and get more and more’ [359 c] (Gk.: πεπλευντικαι: greediness, claiming more than one’s share). This, he says, is what anyone’s nature ‘naturally pursues as good’ [359 c]. Someone who didn’t want to do injustice, and who didn’t want to touch other people’s property, given the opportunity to do so with impunity, would be thought wretched and stupid by everyone aware of the situation, though of course they’d praise him in public, deceiving each other for fear of suffering injustice. Thus, he concludes that no one is ever just willingly, but only because they are constrained by law to be. If self-indulgence could be practiced without fear of punishment, then the tendency for being unjust would prevail.

The story of Gyges remains important for us today because it concerns, among other things, what we can expect humans to do with power over others. In politics, for example, we give power to others – to lawmakers, government officials, and other heads of state -- hoping they will do what is right. If the story is correct, then we had better be careful over how we give power to others. For, according to the story, anyone who gains power without accountability (‘invisibility’) is liable to use it unjustly. Secrecy is a form of invisibility, and for the purposes of power, as effective as a magic ring in perpetrating injustice.

Are we by Nature Egoists?

In many ways Glaucon’s story is compelling. As individuals become richer they are able to satisfy more of their wants. But their wants multiply about as fast as their resources. So the distance between their desires and their ability to satisfy them never quite diminishes. As incomes rise, the aspiration level does too, for, as we get richer, we compare ourselves to a new group of peers who have roughly as much as or more than we do, and we seek what they have. Since there are always people in this world who are richer, we will always have someone to make us miserable, so it seems we can never be truly content. But if given the opportunity to commit injustice with impunity to satisfy our desires, would we seize it? Is injustice rooted in our nature in the way Glaucon claims? This bald thesis appears to fly in the face of many occasions on which people act without any apparent concern for their own interests. Firefighters endanger their lives to rescue people. Nuns spend their lives working among the poor.
The Myth of Gyges and the Possibility of Altruism

People donate anonymously to charity. They donate vital organs to unknown recipients. They build homeless shelters. They volunteer in soup kitchens. The list goes on. Don’t these examples demonstrate that Glaucous’s claim is false? 

Not unless they succeed in refuting the following thesis:

[E] (Psychological egoism (or just ‘egoism’ for short)) Every action is motivated by whatever protects or promotes our self-interest, either because we are indifferent about that of others or because we care more about our own when the two conflict. We do not help others for their own sake, but only as a means of promoting our own self-interest.

[E] is a descriptive theory. It makes a factual claim about human nature: it describes what in fact motivates human behavior. It is a description of psychological facts. It does not prescribe what should motivate human behavior or how humans ought to behave, and hence is not a normative theory. Note, in addition, that [E] does not make any substantive claims about what one’s self-interest is.

As it has been defined, egoism is not necessarily morally objectionable. Morality demands that we be unselfish. How unselfish is a hard question. But arguably, there is nothing morally wrong with caring more about my own well-being than about yours. Egoism becomes morally objectionable when taken to an extreme: if putting my interests above yours requires harming you or violating your rights, I may (depending on the circumstance) be acting unjustly. I will refer to this brand of egoism as amoral egoism. It is this strong form of egoism that Glaucous attributes to human beings in the myth. According to Glaucous, man is by nature unjust. His primary aim is always to help himself, even if that requires harming others. But he cannot always get away with harming others. If he could, he would.

Egoism is contrasted with altruism (disinterested – i.e., non-self-interested -- concern for the welfare others).

---

1Egoism is compatible with action that fails to maximize perceived self-interest, such as weakness of will, since when I act from weakness of will I am still aiming at my self-interest. For example, one may aim to cease their addiction to drugs or alcohol, knowing or believing that it is bad for his health, that it may destroy his marriage, alienate his children and friends, and will in other ways ruin his life. He is weak, though, in that he does not act as he aims.

2To be an egoist, one need not consciously consider one’s self-interest every time one acts; it is sufficient that one’s voluntary behavior conform to this pattern.

3This is why Plato took irrational desires, or failing to recognize ones limits, as the root of injustice.

4Indeed, as Kurt Baier observes [1991 p. 197], as we ordinarily use the term, to call someone an egoist is to ascribe a moral flaw to him. We regard someone as an egoist when he fails to restrain the pursuit of his own good in situations where it conflicts with others, and it is morally required or desirable that he observe that restraint. This is why, in common parlance, we think of egoists as ‘self-centered, inconsiderate, unfeeling, unprincipled, ruthless self-aggrandizers, pursuers of the good things in life whatever the cost to others.’ [1991 p. 197]

5Of course the behavior in question cannot be merely accidental to count as altruism. If I accidentally drop a hundred-dollar-bill on the street without noticing it, it has definitely cost
behavior that benefits others, but in order to establish the contrast with egoism, it must be taken to refer to behavior that benefits others as an end in itself, not as a means for promoting one’s self-interest. It will be useful to construct a chart that distinguishes extreme from moderate forms of egoism, and extreme from moderate forms of altruism, to guide our discussion.

<table>
<thead>
<tr>
<th></th>
<th>Primary Aim</th>
<th>Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egoism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>Extreme (amoral egoism)</td>
<td>Help yourself by harming others</td>
</tr>
<tr>
<td>II</td>
<td>Moderate</td>
<td>Help yourself by doing little or no harm to others</td>
</tr>
<tr>
<td>III</td>
<td>Weak¹</td>
<td>Help yourself by helping others</td>
</tr>
<tr>
<td>Altruism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>Weak²</td>
<td>Help others by helping yourself</td>
</tr>
<tr>
<td>V</td>
<td>Moderate</td>
<td>Help others by doing little or no harm to yourself</td>
</tr>
<tr>
<td>VI</td>
<td>Extreme</td>
<td>Help others by harming yourself</td>
</tr>
</tbody>
</table>

Amoral egoism, being a form of egoism, assumes that every action is motivated by self-interest. Hence, for the examples of helping behavior cited above to succeed in refuting Glaucon’s thesis (amoral egoism), they must (at least) succeed in refuting [E]. Do they?³ Not according to the egoist, for he

---

¹Here it is important to distinguish morally loaded from non-morally loaded interpretations of self-interest. Moral self-interest involves goods received in return for helping behavior that improve one morally, and hence that have beneficial effects on others. Non-moral self interest involves goods received in return for helping behavior (such as material goods, power, fame, fortune, and bodily pleasures) that do not, or do not in any obvious way, have this effect. The literature sometimes refers to non-morally loaded interpretations of self interest as ‘narrow self-interest’. For Plato, the sophists mistakenly identified man’s self-interest with narrow self-interest; they were mistaken, because man’s self interest lay in being just. Plato’s conception of self-interest, based on virtue, was further developed by Aristotle and the Stoics. Since for these philosophers, morality could be reconciled with self-interest, they had ‘a ready explanation to hand of the ease with which moral incentives emerge in the course of psychological development, and of the urgency with which moral considerations present themselves in practical deliberation: these phenomena would reflect the natural concern of humans for their own well being, and the equally natural use of practical reason in the service of this concern’ [Wallace 1998]. My discussion in this essay will not focus on whether morally loaded interpretations of self interest should count as a form of egoism (category III) or altruism (category IV). Perhaps another category in our chart needs to be created between (III) and (IV) for cases, such as these appear to be, where an agent aims at benefiting both himself and others without prioritizing himself over others or others over himself.

²A helpful example of this version of altruism is offered in Rudebusch [2007 p. 81]: ‘The familiar rule for managing a loss of air pressure in an airplane is to put an oxygen mask on oneself before putting masks on those needing assistance. The aim of this self-first rule is to maximize health in general, not to maximize one’s own health’.

³It is, of course, more challenging to refute egoism than to refute amoral egoism. So that is why I am proceeding this way. Obvious difficulties arise for amoral egoism [AE] when we consider
The Myth of Gyges and the Possibility of Altruism

would say that the acts are really not as unselfish as they seem. According to the egoist, human beings are not capable of being unselfish. We may believe ourselves to be noble and self-sacrificing. People may appear to be acting unselfishly and disinterestedly. But that is only an illusion: in reality, every action is motivated by self-interest. Disinterested concern for the welfare of others (IV-VI in our chart) does not exist. Apparent helping behavior is really motivated by such things as:

- the desire for reciprocal helping behavior
- the desire for public recognition
- the desire to receive gratitude
- the hope of heavenly reward

and so is, at root, selfish. Hence, for the egoist, the examples cited above are really instances of (III) in our chart, not (IV-VI).

According to Thomas Hobbes [1651], for any act of apparent altruism, we can find a way to explain away the altruism and replace it with an explanation in terms of more self-centered motives. Consider, for example, charity. We describe people’s behavior as ‘charitable’ when we think they are acting from a concern for others. The American Heritage Dictionary defines ‘charity’ as ‘the love of man for his fellow men; brotherly love’, ‘an act of benevolence, good will, or affection’. Hobbes, however, suggested that ‘charitable behavior’ can be understood in a radically different way: ‘There can be no greater argument to a man, of his own power, than to find himself able not only to accomplish his own desires, but also to assist other men in theirs: and this is the conception wherein consisteth charity.’ Thus charity is a delight one takes in the demonstration of one’s powers: a ‘charitable’ deed is, for its performer, only an opportunity to congratulate himself on the good fortune or cleverness that allows him to be charitable. As Rachels [2003 p. 66] puts it: ‘The “charitable” man is demonstrating to himself, and to the world, that he is more resourceful than others: he can not only take care of himself, he has enough left over for others who are not as capable as he. In other words, he is just showing off his superiority.’

Following Rachels, I will refer to this as the strategy of reinterpreting motives. This is a persuasive – or, rather, seductive - method of reasoning how it is related to amoral normative egoism [ANE]. [AE] is a descriptive theory: it maintains that human beings always pursue their self-interest even if that involves harming others. [ANE] is a normative theory: it maintains that human beings should always pursue their self-interest even if that involves harming others because they are better off that way. [AE], then, would appear to entail [ANE], unless evolution messed up and created us in a way that does not promote our self-interest / is not beneficial to our species. Evolution closes the gap between the descriptive thesis (how we do live) and the normative thesis (how we should live). So, if we have reason to doubt that [ANE] is true, we also have reason to doubt [AE] is true. There are numerous examples in the literature (among them, examples modeled after a prisoner’s dilemma -- see, e.g., Axelrod, Robert [1984]: The Evolution of Cooperation, Basic Books) that give us reason to doubt the truth of [ANE]; hence, they would also give us reason to doubt the truth of [AE].

51
because it appeals to a certain skepticism in us: a suspicion that others are not nearly as noble as they seem. News stories of malice and greed abound to feed these suspicions. Indeed, often we even doubt the purity of our own motives: ‘Is it really something fine and noble I desire, or to be thought well of by others and enjoy the pleasures of a good conscience?’ It is a well-known fact that people tend to conceal their true motives from themselves, especially when they are disreputable. ‘It is only natural that we would interpret our actions in a way that is flattering to us.’ [Rachels 2003 p. 66]

The lure of egoism can also be traced to a less flattering reason: denying the possibility of altruism ‘provides a convenient excuse for selfish behavior. If “everybody is like that” – if everybody must be like that – we need not feel guilty about our own self-interested behavior or try to change it.’

Thus, for every apparent case of altruistic behavior, the egoist can argue, with some plausibility, that the true motive is really selfish, appearance to the contrary. Philanthropic acts are really motivated by the desire to receive gratitude. Acts of self-sacrifice are really motivated by the desire for recognition. ‘Some of us might find any substituted egoistic explanations more plausible than non-egoistic ones, because we already believe that deep down we are all egoists.’ [Baier 2001 p. 199]

Despite the many unmasking explanations which Nietzsche, Marx and Freud have accustomed us to, however, we would not be warranted in concluding that the egoistic explanations are more satisfactory/plausible than non-egoistic ones. In other words, the method of reinterpretting motives ‘is not a conclusive method of reasoning’. [Rachels 2003 p. 67] It only shows that it is possible to interpret motives egoistically; it does not show that the egoistic explanations are correct or even more plausible than their competitors. Admittedly, ‘a clever egoist can often invent a plausible underlying egoistic explanation of apparently non-egoistic behavior, just as a dissembling egoist can substitute an invented, nobler explanation for an egoistic one’. [Baier 2001 p. 199] But insisting that there must be an egoistic motive does not make it the operative one. The strategy does not show that egoism is true.

One may object: truth is too much to ask for here. At best, what we can and should demand from the egoist are reasons to believe that his theory offers a better explanation of human behavior than its competitors. After all, egoism is a psychological/explanatory theory, and, as such, demands empirical evidence to back it up. What reasons, then, can the egoist offer to show that his theory best explains human behavior?

Some philosophers, psychologists, and economists, have turned to evolutionary biology to supply the needed evidence. Biologists noticed that animals frequently help each other; that monkeys, for example, scratch and meticulously pluck fleas from the backs of fellow monkeys. They asked the perfectly good question of how such behavior could have emerged. The answer proffered was that it has a functional or adaptive purpose: it triggers reciprocal helping behavior from the animal helped, or from others witnessing the event.

Generalizing from animal behavior to human behavior, egoists infer that, likewise, apparent helping behavior in the human species has an adaptive or functional purpose: to trigger reciprocal helping behavior. ‘A public benefaction is just plain good business advertising; the “Golden Rule” is no more than an eminently sound success formula’, etc. In other words, according to the egoist, apparent helping behavior is an instance of (a), not (b):

(a) ‘I’ll scratch your back and you scratch mine.’ (egoism)
(b) ‘I’ll scratch your back and never mind mine.’ (altruism)

(b) involves disinterested concern for the welfare of others, which, the egoist claims, does not exist.

However, even if these claims offer some support to the claim that some forms of apparent helping behavior are ultimately self-serving, they fail to address, or at least appear to fail to address, other important counterexamples to the egoist’s thesis: cases where people help each other when there are no bystanders to witness the event, and hence no apparent means of initiating reciprocal helping behavior; cases, in other words, where we appear to help others for its own sake, and not for the sake of helping ourselves. Consider the person who donates anonymously to charity, or who tips a waiter he knows he will never see again, or who gives his sandwich to a homeless person; all this when there are no bystanders to witness the action. Examples like this can also be found in the animal kingdom. Dogs often adopt orphaned cats, squirrels, ducks and even tigers. Dolphins support sick or injured animals, swimming under them for hours at a time and pushing them to the surface so they can breathe. I stumbled upon the following interesting example concerning a chimpanzee and a human on the internet:

A researcher was following a group of chimpanzees in the jungle. After some hours, he found that he had forgotten his lunch back at the research station. This researcher then proceeded to try to knock down fruit from a tree some distance from where the group of chimpanzees sat eating their mid-day meal. After some time of unsuccessful attempts to acquire fruit, a young male from the group collected some fruits from a tree and climbed down toward the researcher. The chimpanzee then proceeded to approach the researcher and leave the fruit for the researcher.

Don’t these examples establish the falsity of egoism once and for all? What can the psychological egoist say in his defense? It seems he has only one possible rejoinder. Assuming the doers are not expecting to be rewarded in heaven for their good deeds. If they believe in an afterlife, then they may regard altruistic acts as ‘assuring them a place in heaven’, and hence as a means of benefiting themselves.

---

2http://scientificcuriosity.blogspot.com/2006/12/example-of-altruism.html
3Assuming the doers are not expecting to be rewarded in heaven for their good deeds. If they believe in an afterlife, then they may regard altruistic acts as ‘assuring them a place in heaven’, and hence as a means of benefiting themselves.
An Anthology of Philosophical Studies

hence belong in category III in our chart), because one engages in them to feel good: to experience the warm glow of self-congratulation, or to avoid feelings of guilt. ‘Unselfish’ actions produce pleasure in the door -- a sense of self-satisfaction -- and this, the egoist might say, is why they really performed them in these cases. Thus we have what I will call the ‘feel good’ argument.

<table>
<thead>
<tr>
<th>The feel good argument</th>
</tr>
</thead>
<tbody>
<tr>
<td>i. When people help others in situations where there are no bystanders to witness the event, they at least feel good about their helping behavior.</td>
</tr>
<tr>
<td>ii. Therefore, the reason they help others in these situations is to make themselves feel good, not because they genuinely care about others, or from disinterested a concern for the welfare of others.</td>
</tr>
<tr>
<td>iii. Therefore, even in these cases, they are egoists, since the ‘zero benefit’ criterion of altruism is violated: they help others, not as an end in itself, but for the sake of helping themselves.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Argument that acting from pity is self-serving</th>
</tr>
</thead>
<tbody>
<tr>
<td>i. When, e.g., you touch fire, pain is experienced, but it acts as a warning, a protective agent, to prevent further pain (damage from fire). As a result of the pain experienced, the hand is jerked back. Therefore, the pain has provoked the physical action of retracting one’s arm. In the future, that individual will be well aware of what touching the flame will result in, and will take precautions.</td>
</tr>
<tr>
<td>ii. Similarly, an individual may have experienced guilt, a bad feeling, provoked by an external source. After experiencing guilt, the individual develops an emotion that would later cause him to take action in order to prevent what will follow. That emotion is pity. The pity felt is not a good feeling, but it acts as a warning, a protective agent, to prevent further mental pain (guilt).</td>
</tr>
<tr>
<td>iii. Just as the pain experienced from the flame acts as a warning that action must be taken to prevent further physical pain (damage from fire), so too pity acts as a warning that action must be taken to prevent further mental pain (guilt). The flame would be analogous to, say, a homeless man asking for a little help, which invokes pity.</td>
</tr>
<tr>
<td>iv. Action is then taken to stop the bad feeling, and prevent a worse feeling from occurring. This is the seemingly good deed, where the individual helps the man, thereby preventing further pain: the experience of guilt. Pity induces precaution (‘good deeds’), to prevent a bad feeling (guilt), and is therefore self-serving.</td>
</tr>
</tbody>
</table>

There is one obvious problem with this argument. Of course, it’s true that when people behave ‘altruistically’ in the cases mentioned, they often feel good about it. But it doesn’t follow that their motive for acting ‘altruistically’ is to feel good, as premise (ii) claims. Though good feelings may be a product of helping behavior, the desire to have them is not necessarily the cause of the behavior. The argument commits the logical fallacy of assuming that correlation implies causation. Suppose, for example, you encounter a homeless person on the subway, and you give him a dollar. As he walks away, you may feel good about what you just did. But it doesn’t follow that you did it because you wanted to make yourself feel good. You might have done it simply because you pitied (or empathized with) him. That, by itself, may be what
caused you to act the way you did, and the good feeling a product of the action. Unless one can show that acting from pity is self-serving or self-regarding, we do not have reason to believe premise (ii).

One might try to show that acting from pity is self-serving by analogy.

Critique

One question we need to ask is: Is acting from pity ‘self-serving’ in a way that should cause us to classify seeming altruistic deeds as egoistic? [In other words, does the conclusion of the feel good argument follow from its premises?] Is it self-serving in the way that, say, cutting someone on line is self-serving? Cases involving self-sacrifice bring out the absurdity of referring to helping behavior that has no payoff beyond generating good feelings in the doer as self-serving. Suppose a soldier risks his life by entering a line of fire to rescue a wounded soldier. It does not seem that the soldier is pursuing his perceived self-interest. It is plausible that, if asked, the soldier would say that he felt it was his duty to save the life of the wounded soldier. He would deny as ridiculous the claim that he acted to make himself feel good, even if he did feel good about the deed afterward. (Of course, the most difficult cases for the egoist are extreme cases of self-sacrifice, such as that exemplified by saints, martyrs, military heroes, patriots, and others who have offered their life for a cause. However, the egoist can weaken his theory to deal with these cases, as follows: ‘All actions other than those where one offers one’s life to help others aim at one’s self interest’, whence the challenge to refute egoism remains.)

In keeping with the feel-good argument, the egoist might insist that the soldier is deceiving himself about his true motives: he did what he did because he could not bear to live with himself afterwards if he failed to do so. He has a better life by avoiding years of guilt.

To see the inadequacy of this reply, contrast the case where a soldier enters a line of fire to save someone’s life, with that of a soldier who pushes a fellow soldier into a line of fire to save himself. We ordinarily think there is a significant difference between these two cases. We think the former is acting unselfishly whereas the latter is acting selfishly. According to the feel good argument, however, they would both be acting selfishly. This is so counterintuitive that some would regard it as a reductio of the feel good argument.

We need, however, to be wary of drawing this conclusion too quickly. Of course it is true that not everyone derives pleasure from helping others or pain from failing to help them. From this it is tempting to conclude that:

(*) Any instance of helping behavior is altruistic if and only if it generates pleasure in the doer.

After all, why should it make you feel good to contribute money to support a homeless shelter, when you could spend it on yourself instead? The answer
seems to be, as Rachels proposes [2003 p. 70] that ‘you are the kind of person who genuinely cares about other people. If you do not genuinely care about them, then giving the money will seem like a waste, not a source of satisfaction. You’ll feel more like a sucker than a saint’.

True, for the egoist, benefiting others has an instrumental value; it is not an end in itself; it is a means toward a further end: benefiting oneself. So we shouldn’t expect him to derive pleasure from helping others, and pain from harming them. Rather, we should expect him to be indifferent to the pleasure or pain his actions have on others. We should expect him to feel good only after his goal has been achieved, toward which he treats helping others as a means. On the other hand, we should expect the altruist to derive pleasure from helping others, and pain from harming them, because he treats helping behavior as an end in itself. As Aristotle points out in the Nicomachean Ethics [1099a15], ‘No one would call a person just if he failed to delight in acting justly, nor open-handed if he failed to delight in open-handed action’. To this we may add: ‘nor an altruist if he failed to delight in acts that benefit others.’ If this is so, then the fact that people feel good about helping others, far from lending support to egoism, can be taken as a sign that their acts are genuinely directed toward helping others as an end in itself.

We must beware, however, of concluding that for any given case of helping behavior, deriving pleasure from the act by itself suffices (and hence that pleasure is a conclusive sign) to classify the act as altruistic (as Rachels [2003] and Feinberg [2008] do). In Feinberg’s words: ‘The presence of pleasure as a by-product of an action … provides rather conclusive proof that the act was unselfish … the fact that we get pleasure from a particular action presupposes that we desired something else — something other than our own pleasure — as an end in itself and not merely as a means to our own pleasant state of mind.’ [2008 p. 523] This conclusion seems too strong, for surely we can imagine, and very likely can find, people who engage in helping behavior merely to make themselves feel good. After all, people engage in other acts (for example, sex) for the sake of pleasure, so why not helping behavior? Suppose, for example, that Smith is generally a benevolent person. He takes great pleasure from helping others. But he discovers that on his trips to certain countries that happen to hate foreigners, he is deprived of this ‘reward’. Waiters and cab drivers generally do not thank him when he tips them. People are rude when he tries to strike up a friendly conversation with them. He is not thanked when he gives up his seat on a train or bus. Even homeless people do not smile when he gives them his pocket change. They accept his help, but view it more as an obligation than a display of generosity. As a result, Smith ceases his benevolent behavior in these countries, but back home he returns to his usual pattern of helping others. It would then seem wrongheaded to classify Smith as an altruist with regard to his helping behavior back home, for he appears to engage in it merely for the good feelings that it brings him.


\[2\] For (at least) this reason, the fact that we are wired to join groups with which we cooperate to accomplish tasks that we would not be able to do alone does not conclusively establish
The Myth of Gyges and the Possibility of Altruism

Now suppose Jones also takes pleasure in helping others, but unlike Smith, his helping behavior is consistent. He has traveled to the same countries as Smith. He too, like Smith, is never ‘rewarded’ with kind words or a smile in those countries. On some occasions when he has helped others, he has been laughed at; on others even cursed at. However, unlike Smith, he persists in his helping behavior. Then it appears that Jones really does treat helping others as an end in itself. The point is, we cannot determine whether an act is altruistic or egoistic by inspecting it on its own. We need to examine the overall habits of the agent, to observe his behavior more generally, before we can classify the act one way or the other. The agent’s motives cannot be ‘read off’ isolated acts.

Plato’s description of the ‘completely just’ man in the Republic, Book II, shortly after his presentation of the Myth, is helpful here.

[He is] simple and noble and … doesn’t want to be believed to be good but to be so. We must take away his reputation, for a reputation for justice would bring him honor and rewards, so that it wouldn’t be clear whether he is just for the sake of justice itself or for the sake of those honors and rewards. We must strip him of everything but justice and make his situation the opposite of an unjust person’s. Though he does no injustice, he must have the greatest reputation for it, so that his justice may be tested full-strength ... [Republic, 361 b-c]

The idea is that, if under these conditions one perseveres in his justice, then we can conclude that he really does love justice for its own sake. I am suggesting the same apply to our altruist. If he persists in his altruism even in situations where he is scorned and persecuted, then we can conclude that he treats his helping behavior as an end in itself.

In conclusion, to show that altruism is real, I submit, we need to find examples that fit the description of altruism I have offered above. It is not difficult to find historical figures that fit the bill. I leave as an exercise to the reader.

References


Altruism. Granted, in groups we can obtain all the things we need for survival, security, and increased longevity so that we improve the likelihood that our children will survive. Granted, alone we do not have the speed and power and weapons of other natural predators but in groups with our intelligence and use of tools, we can master any prey. And granted, we form emotional bonds based upon such affiliations. Indeed, neuroscience has established that our brains are full of neurons that react to what others experience as if we are experiencing it: we generally suffer when we witness others suffer. But as I just noted, sharing someone’s suffering, and acting to reduce his suffering, is not necessarily a conclusive sign that altruism exists, for you may be merely acting to reduce your own suffering. So the evidence from group behavior is not conclusive evidence for altruism.


CHAPTER SIX

Aristotle’s Theory of Responsibility and Situationism

Panos Kapetanakis

Situationist social psychology presents an unsettling argument against freedom and responsibility (Nahmias, 2010: 353-54):

(1) Freedom and responsibility require that one’s actions properly derive from reason-responsive states or processes, like decisions, reasons or character traits.

(2) Studies in social psychology suggest that our actions are often caused by situational factors of which we are unaware or rationalize post-hoc.

(3) So, science shows that we are less free and responsible than we think.

The argument suggests that person-based factors are not the bulwark against the causal influence of situations and luck. Situationism is not a form of causal determinism, though. Skeptics seek to show that responsibility does not originate in our psychology properly. Aristotle’s account is the paradigm and origin of this idea: responsibility lies in our character because it is reason-responsive. Like situationism, his approach is un-metaphysical. So, psychology is the scene for the dialectic between Aristotle and skeptics.

However, the dialectic is not based on a common paradigm of psychology. On the one hand, Aristotle’s account is couched in terms of folk psychology. On the other hand, the scientifically-minded skeptics distrust folk psychology, reasons-explanations and they take a sharp person / situation dichotomy for granted. Part of my work here is to examine these assumptions.

In section 2, I present the skeptical argument in detail. In section 3, I reconstruct Aristotle’s account of responsibility. In section 4, I examine the argument and suggest that it is not as compelling as friends of situationism claim. Aristotle’s account is a valuable source for spelling out views of

1 Thanks to Neera Badhwar, Konstantina Dizelou, Christos Kyriacou and Michael Michael for helpful feedback, comments, and discussion. Special thanks to my Ph.D supervisor Dory Scaltsas who brought this problem to my attention and patiently guided me through the exegetical and philosophical puzzles of Aristotle’s work on responsibility.

2 Some commentators, ancient and contemporary, have tried to trace Aristotle’s assumptions about freedom and determinism. For example, Alexander Aphrodisias, Hughes (2001: 137-142) and Broadie (1991: 149-159) attribute to Aristotle some form of libertarianism, whereas Meyer (1998) thinks that Aristotle would subscribe to Stoic compatibilism.
responsibility which are safe from the situationist challenge.

**Situationism: Character and Responsibility**

*Skepticism about Character*

Let us see some notable situationist experiments. A study by Isen and Levin (1972) suggests that subjects who had just found a dime in a payphone were fourteen times more likely to help a woman pick up her dropped folder than subjects who did not find a dime. In Darley and Batson’s study (1973) subjects who were in a hurry were six times less likely to help a man slumped in a doorway in obvious distress than subjects who were not in a hurry. Darley & Latané (1968) demonstrate that the probability of helping is inversely proportional to the number of bystanders in an emergency situation—this phenomenon is called bystander effects. Also, in his famous *obedience to authority* study Stanley Milgram (1974) reports that two-thirds of the participants would consistently administer a 450-volt electric shock to someone if an authority figure asked them to do so.

The situationist interpretation of these findings is that in any helping situation the likelihood of offering help does not depend on a character trait (e.g. benevolence) but on situational factors (e.g. the hurry factor). Hence, a difference in two persons’ helping behaviour is better explained by some difference in their situations. Finding a dime in a telephone booth is a mood boost while a bad day at the office is not.

A further situationist conclusion is that character traits are not cross-situationally consistent or robust: a person may be repeatedly helpful with her friends but not repeatedly helpful with her colleagues and neighbours. If character traits were robust, then they would be ‘reliably manifested in trait-relevant behavior across a diversity of trait-relevant eliciting conditions’ (Doris, 2002: 22). Situationists claim that any observed behavioural consistency is primarily context-specific.

A number of philosophers argue that these findings suggest the death of character and virtue ethics.¹ Since character traits do not lead to ‘patterns of behavior expected on broad trait categories like “introverted”, “compassionate”, or “honest”’ (Doris, 2002: 25)—i.e. the so-called virtues—then there is nothing substantive for virtue ethics to be about. Friends of virtue theory contend that the situationist experiments are not relevant to the conception of character found in Aristotelian and Neo-Aristotelian writings.² The debate is ongoing.

---

¹See Harman (1999) and Doris (2002).
Aristotle’s Theory of Responsibility and Situationism

The Situationist Argument (1)\(^1\)

Let us see the bearing of these findings for moral responsibility. I said before that the focus of my discussion will be premise (1) of the situationist argument. Doris (2002) and Nahmias (2001) think that responsibility requires identification with, or acceptance of, one’s determinative motives or reasons for acting.\(^2\)

To the extent an actor identifies with the determinative motive of a behaviour she performs, she is responsible for that behaviour; to the extent she does not so identify, she is not. (Doris, 2002: 140)

When we report on the reasons for our attitudes or actions, we usually mean both that those reasons have influenced us and that they justify our attitudes or actions—that is, we see the reasons as causal influences and, usually, we accept them as legitimate influences. So, to the extent we actually act on these reasons, we are acting on motivations we identify with, and hence acting of our own free will. (Nahmias, 2001: III.2; my emphasis)

Doris (2002), Nahmias (2001) and (Nelkin (2005) argue that identification requires knowledge of our determinative motives and normative competence—the ability to distinguish right from wrong and to control behaviour reflectively.

Premise (2) suggests that the empirical conditions of identification do not obtain. First, situational factors, like bystander effects, go unrecognized. Some participants report causes that had little or no influence on their actions; others rationalize their actions post-hoc. Second, Darley & Bateson (1973: 108) report that the hurry factor often distorts moral perception: ‘[Some participants] did not perceive the scene…as an occasion for ethical decision’. Other hurrying participants perceived the need to help but failed to act in accord with that evaluation—normative competence is impaired.

Let me restate their argument: (1) responsibility requires identification, which further requires self-knowledge and normative competence; (2) situationist experiments suggest that we often fail to identify with our actions because our determinative motives are factors of which we are unaware or would not accept; also, our normative competence is often impaired by situational influences. (3) Hence, we are less responsible and free than we think.

The Characterological Argument (2)

This second version of the situationist argument denies that responsibility lies in robust character traits. One attempt is Doris’ (2002: 129):

\(^1\)The bearing of situationism on moral responsibility is primarily discussed by Nahmias (2001; 2007; 2010), Doris (2002) and Nelkin (2005).

\(^2\)This view is identificationism. Frankfurt (2003) argues that our freedom and responsibility lie in our capacity for reflective acceptance of our motivational states by our higher-order volition. Doris and Nahmias use identificationism as a cogent account of responsibility.
Deep responsibility assessment must be properly psychological assessment, and (2) properly psychological assessment involves character assessment, then (3) skepticism about character precludes deep responsibility assessment.

For Doris it is an error to blame a selfish act thinking that one has a robust trait, selfishness. Situationism suggests that we do not have robust traits. So, these responsibility attributions are unfounded. Another way to state the argument is as follows (Nelkin, 2005: 189; Nahmias; 2001: 19-20):

(1*) Responsibility requires that we identify with robust character traits.
(2*) Situationist experiments suggest that we lack robust character traits.
(3*) So, we are less free and responsible than we think.

The idea here is that responsibility requires acting on a trait-ideal that we embrace. But, according to Nahmias (2007: 7), situationism precludes this: ‘To the extent that the dispositional traits we identify as our principles do not in fact correspond with consistent behavior, we are identifying ourselves with constructed concepts rather than actual motivational states. This limits the influence of our principles on our actions.’

Let us take stock: responsibility requires identification with principles expressed by robust traits; situationism suggests that character traits are not robust—our character ‘disappears’ due to situational influences; hence, character traits cannot serve as the objects of our identifications. Hence, our responsibility is more limited than we think.

We now have a picture of why skeptics take the situationist experiments to imply a direct threat to moral responsibility. Let us move now to Aristotle’s account. Apart from being the first to have mapped this conceptual territory, Aristotle offered a valuable analysis of the conditions of moral responsibility and a compelling defence of the practice.

Aristotle on Responsibility

The experimental evidence from situationism is unsettling for any ethics that takes responsibility seriously, like Aristotle’s. His explicit concern with responsibility in his Ethics is practical: he seeks to analyze culpability and excusing conditions which are at the heart of ethical and jurisprudential thinking, and Athenian law. Here is Aristotle’s main idea:

---

1A number of social psychologists call this the fundamental attribution error. This is the tendency to cite robust character traits in action explanation and to underestimate the influence of situations; see Jones & Harris (1967); Ross (1977).

2This is most clearly stated by Nahmias (2007: 7)—from whom I borrow the metaphor of disappearance—and Nelkin (2005: 189).

---

62
Since, then, excellence [i.e. virtue] has to do with affections and actions, and in response to what people do that is voluntary we praise and censure them, whereas in response to what is counter-voluntary we feel sympathy for them, and sometimes even pity, those inquiring into the subject of excellence [i.e. virtue] must presumably determine the boundaries of the voluntary and the counter-voluntary; and to do so it is also useful for those framing laws, when it comes to fixing honours and methods of forcible correction. (NE 1109b30-35)

According to the above, an action is subject to praise and blame if and only if it is voluntary. Aristotle states further: an act done under force or ignorance is involuntary (NE 1110a1). Let us unpack his definition in the positive. First, a person acts voluntarily just in case: (a) the moving principle of the act originates in the person; (b) the person has knowledge of facts related to the act and its foreseeable consequences. For instance, serving a drink that one knows is poisonous amounts to killing one’s guest voluntarily. Let us sum Aristotle’s definition of voluntariness:

(V) A does X = act voluntarily if and only if (a) the moving principle of X is in A, and (b) A has knowledge of facts about X.

Aristotle considers a number of excusing conditions to (V). We already saw that force and ignorance are excusing conditions. But not all forms of ignorance are excused. First, Aristotle argues that feeling no regret for doing harm due to factual ignorance is blameworthy (NE 1110b20-25). Second, causing harm in ignorance—e.g. due to drunkenness—is not an excuse because it is willful ignorance. In contrast, causing harm because of ignorance is an excusing condition—as stated in (V) above. Third, ignorance of norms—e.g. that stealing is wrong—is not an excuse—everyone should know such universals (NE 1110a35).

Then Aristotle discusses whether compulsion or duress is an excuse. He says that acts under duress are mixed acts. They are voluntary because we cause them but involuntary because no good person would choose them for what they are. In general, mixed acts deserve pity, not blame, because they are done in circumstances that ‘overstrain human nature’ (NE 1110a25), like the sea-captain who in the midst of a storm throws away cargo to save his ship and crew (NE 1110a12-20). However, compulsion is not always a mitigating factor: Being forced to slay one’s mother, like Euripides’ Alkmaeon did, is not acceptable. Aristotle argues that Alkmaeon should have faced death instead (NE 1110a28). However, he offers no clear condition for separating forgivable from unforgivable mixed acts.

Lastly, Aristotle rejects the Socratic claim that wrongdoers act involuntarily because they ignore the real good (Protagoras 352a-c; Gorgias 468b; Meno 77e-78b). He agrees with Socrates that all voluntary action aims at some good
An Anthology of Philosophical Studies

(NE 1094a1-3) and that wrongdoers act under the guise of the good. So, they mistake the apparent for the real good. However, they are culpable for this mistake because their character influences what things appear good to them (NE 1114a311-b26). How does Aristotle support this claim? Here is his controversial argument: Our character formation depends on us (NE 1113b6-7) because moral dispositions result from doing like acts voluntarily—this is his principle of habituation (NE 1103b22-3). Hence:

... [W]e are ourselves responsible for having become this sort of person, by living slackly, and for being unjust or self-indulgent, in the first case by treating people badly, in the second by passing our time in drinking and that sort of thing; for it is the sort of activity we display in each kind of thing that gives us the corresponding character. (NE 1114a5-9)

But what if one is not aware that moral dispositions result from doing like acts? Aristotle argues that to ignore this is a sign of a thoroughly senseless person (NE 1114a9). Unfortunately, he does not provide support for this claim although this is a crucial step in his argument. One may also complain that habituation occurs at too early an age to be a product of knowledge, reflection and choice.

However, we should not overlook that habituation of the affective self is just the first step of character formation.¹ Aristotelian character education does not aim to inculcate habits of action. Nor does it end at an early age. Rather, moral dispositions are formed gradually with the aid of practical reason since, according to Aristotle, our affective self ‘is capable of listening to it [i.e. reason] and obeying it’ (NE 1102b32). Hence, the more our rational and cognitive capacities develop, the more the acts that generate our dispositions get under reflective control. As Aristotle says, our virtues and vices are up to us (NE 1113b6-7).

Why are they up to us exactly? His answer is somewhat technical: because we act on objects of deliberation and decision, i.e. on things that depend on us (NE 1111b30). For example, we do neither deliberate about, nor choose to find, ‘a cache of treasure’ (NE 1112a28) although we may wish we found one. So, it is up to us whether we exercise our dispositions on good or bad objects. Since dispositions get gradually settled, they are in principle revisable because character-virtue / vice is ‘disposition that issues in decisions’ (NE 1106b36). In other words, our dispositions partake in reason whose role is to evaluate ends and discover good ones. So, our dispositions are under the evaluative control of reason.

I sketched how Aristotle defends his view that we develop and acquire vices and virtues voluntarily. The upshot of this sketch is that our responsibility lies in our reason-responsive dispositions which are ultimately up to us. I have argued that, for Aristotle, character is not fate—i.e. there is an ongoing shaping

¹There is increasing agreement in Aristotelian exegesis that character formation is ongoing and aided by reason; see Burnyeat (1980) and Sherman (1991).
and re-shaping of our characters. Also, I have assumed that, for Aristotle, reason is not a slave of the passions—practical reason can discover the good ends. The limited space here disallows a thorough exegesis of Aristotle but I think my reading is defensible.\(^1\) Besides, this is what Aristotle should say in order to establish that we are culpable for our moral ignorance and our characters.

Aristotle and Situationism

Culpability and the Moral Agent

We saw that premise (1) of the situationist argument states psychological requirements about responsibility and premise (2) puts them to the situationist test. My concern here is to see whether the requirements that skeptics target in (1) are necessary and / or sufficient for responsibility or not. If they are not, then the argument is not convincing. But, first, we need to understand what goes under the rubric ‘responsibility’. It appears that several different projects go under this term: conditions of culpability, conditions for being a moral agent, the concept responsibility, and perhaps more.\(^2\) So, we need first to ask which aspect of responsibility premise (1) is about.

My suggestion is that the conditions which distinguish moral agents from non-moral agents are different from those that distinguish culpable agents from non culpable agents within the class of moral agents.\(^3\) The former conditions make someone a moral agent. The latter conditions make a moral agent apt for attributions of culpability, not for attributions of moral agency. When someone fails to meet the conditions for moral agency, we exempt them from the moral community. When moral agents meet no culpability condition, we excuse them. This is not an exhaustive distinction but a way to disentangle two different questions.

Situationist skeptics seem to conflate these two questions. They view identification—and its components—as a culpability condition and argue that we are less culpable than we think. This would be no surprise because wrongdoers often do not identify with their determinative motives, as when someone kills in a fit of rage may not identify. On the identification condition, even heinous acts should be excused. But we want to say that, unless they are not moral agents, wrongdoers are culpable when they act voluntarily.

\(^1\) Aristotle does not preclude character reform. He only says that those who have done irreversible damage to their characters can no longer restore their moral health—this is like trying to restore a stone which we have just let go (NE 1114a15-20). For the gloss that reason can deliberate on, and evaluate, ends see Sorabji, R. (1980); Wiggins (1980). Aristotle’s psychology allows character reform because when reason discovers some good, wish follows and motivates automatically; see (DA 433a22-5); Cooper (1999). Hence, we can start acting contrary to vicious desires and generate better dispositions.

\(^2\) A useful discussion about the components of a comprehensive theory of responsibility is Eschleman (2009).

\(^3\) My distinction follows Strawson (1962) who introduced the distinction between excusing and exempting conditions.
Lack of identification (e.g. due to insanity) would be an exempting condition, not an excusing condition. Someone who seriously lacks reflective control of one’s motives, and the ability to distinguish right from wrong, cannot really be a moral agent. But that would exempt someone from moral agency; it would not excuse a moral agent from culpability.

Without this distinction, skepticism will appear plausible because identification requires that we be aware, rational, self-controlled decision-makers. It is no wonder then that very few acts would be culpable. But they should not overlook that these capacities are necessary for being a moral agent, not for culpability.

**Aristotle: Moral Agents as Characters**

Aristotle’s culpability condition is voluntariness. Hence, voluntariness is not debatable on the grounds of (1). However, skeptics may use (1) against Aristotle’s view of what makes someone a moral agent—i.e. one’s reason-responsive dispositions or character. I examine this challenge in the last section. For the moment, skeptics could object to Aristotle’s culpability condition, not via (1), but on epistemic grounds. Let us examine this problem.

Aristotle says that one acts voluntarily just in case one has factual and normative knowledge. As it happens, one’s factual and normative knowledge is a correct description of reality only if it is a true belief. In the moral case, things get tricky. While we should know norms, like N, we often fail to perceive a situation requiring N, as an instance of N. Someone who fails to perceive an instance of N as N, seems to act under a different description, say P, without realizing it fully.

Situationism suggests that situational factors may limit or impair our normative competence. So, skeptics may contend that under such circumstances, wrongdoers are not culpable because they do not act knowingly; hence, they do not act voluntarily. For example, they may know that selfishness is a vice but fail to perceive a situation requiring to share—e.g. due to being in a hurry. In such a case, being in a hurry would be an excuse.

For Aristotle this is not an excuse. While wrongdoers are ignorant of what a situation requires, some vice makes bad things appear good to them; that is how they become ignorant. Selfish people are ignorant of others’ needs but they act in ignorance because they are too much absorbed with themselves, careless or indifferent. Aristotle showed above how we let ourselves become ignorant willfully. So, wrongdoers should not be ignorant because their dispositions—their moral lens—are up to them. We are moral agents and ultimately responsible for our moral ignorance—which is ultimately caused by vices of our character—because our character is up to us. In sum, this is Aristotle’s condition for moral agency:

(R) One is a moral agent in virtue of one’s character which is the basis of one’s responsibility.
Most commentators agree that Aristotelian practical reason evaluates ends and discovers the good. Also, reason can motivate action through wish (i.e. rational desire) against vicious desires; wish is an intellectual appetite for the good \((DA \text{ } 433a22-5)\). Given this view of reason and its relation to character states, it follows that our dispositions are not mindless habits of action. Hence, we are not automata. We act and feel for reasons. Our views of the good depend on our characters and are revisable by reason.

**Character in Intentional Explanation**

One basic assumption of skeptics is that ‘ethical reflection should be predicated on a moral psychology bearing a recognizable resemblance to actual human psychologies’ (Doris, 2002: 112). This useful principle is called psychological realism (Flanagan, 1991: 32-38). Surely, scientific psychology constrains unduly optimism in our autonomy. But the amount of realism depends on what we will test scientifically. That is, if we adopt a very demanding requirement for responsibility—say, to become moral saints and abandon our other commitments—then we will consistently fail it. Then moral responsibility will become chimerical—some normative practices \((ought)\) are not psychologically realizable \((can)\).

Aristotle’s \((R)\) is not demanding and facilitates our actual practice of responsibility attributions. On the other hand, skeptics use demanding requirements for both moral agency and culpability: we must be aware, rational, self-controlled decision-makers. We saw that these are unreasonable demands on culpability and that is why they lead to skepticism. But they are reasonable demands on moral agency. We must have these capacities; otherwise we will be morally adrift. Skeptics may object that we are creatures of habit and automaticity; a large part of our behaviour takes place automatically, habitually or pre-consciously; so, we are not as autonomous as we assume. However, this is not right because automatic behaviour is controllable (e.g. driving a car) or stems from previous decisions and reasons.

Skeptics may also object from premise (2): there is not much psychological realism in the idea that we act on our own reasons because proper moral agency requires more reasons-responsiveness than our psychology allows; but situationism suggests that often our reasons for acting are unrelated to the real causes of our behaviour—i.e. the situational variables. This might mean (a) that we are not acting for reasons at all but under stimulus control like automata. Or it might mean (b) that the reasons for which we think that we act are not the reasons why we acted.

Gloss (a) is strikingly odd. How could situations get a causal grip on our behaviour unless by becoming normative or motivating reasons? Situations cannot be the only determining causes because there have to be some features in situations which appear somehow favorable or obligatory to us. Hence, situational variables could affect our behaviour only in virtue of person-based factors, like our appraisal of situations as reasons for acting.

Gloss (b) is reminiscent of the problem of ignorance. Skeptics seem to suggest that we act for reasons but we are not aware of them. These are
phenomena like self-deception or subconscious motivation. But skeptics cannot ignore that, still, these reasons are attributable to us. As long as one accepts that all voluntary action is done for reasons, the chain goes back to agents who have certain motivational and cognitive traits. Hence, character seems always to be a part of the explanation of action.

This substantiates Aristotle’s (R) because it necessitates character in the explanation of intentional action. Yet, a hard test for (R) is premise (1*) and the scientific evidence from (2*): character-based responsibility requires robust character traits; our traits tend to disappear under situational influences; hence, our character cannot be the basis of our responsibility.

However, situationists falsely attribute to Aristotle a view of character as consisting of robust traits. If Aristotle held such a view, then (2*) would undermine it. Principle (R) does not require robust character traits. First, it requires that actions spring from states within us which are in principle reason-responsive. We already saw that moral dispositions have that nature. Second, it requires that both automatic and deliberate actions are attributable to character. Aristotle’s approach secures that, too: our moral dispositions become habitual, automatic but reason-responsive. Hence, even less conscious agency is still responsible. Third, it requires that these states are somehow involved in all voluntary action for reasons. Dispositions fill this role because they are motivational patterns of seeing which are sensitive to reasons. Fourth, responsibility requires rational self-control. Our dispositions are reason-responsive and allow us to do otherwise. The account I sketched satisfies key platitudes about responsibility: ownership or origin of action, control or the ability to do otherwise, self-determination and freedom.

Concluding Remarks

Situationism may bring to mind the problem of moral luck. Luck can make a difference in what agents are blamed for, whether they become good, and whether they can sustain their goodness (Nagel, 1979). The situationist commentary on the atrocities at Auschwitz and Dachau, the Rwandan genocide, the infamous Milgram’s and Stanford Prison experiments, sounds unsettling (Doris, 2002). Who were these people? Presumably, most of them were not psychopaths or sadists but ordinary people. For some of them the situational weight was heavier than their characters could lift. Others went with the flow. Aristotle would say that maybe they were ethically unprepared or untutored. As I see it, his answer to these disheartening views is that, in principle, it is open to us to improve. Luck does not disallow this possibility. Situationism can alert us of the moral perils of situations and inform our efforts to improve ourselves, our institutions and policies so that we make our world a better place. It cannot preclude that this is up to us.
References


C H A P T E R  SEVEN

Geometry without Space:
Ancient Greek Mathematical Thought and
Contemporary Consequences

Duane J. Lacey

In the 4th century BC Euclid composed his well-known text the Elements, introducing to the world the first known example of an axiomatic treatise and system that was also considered to be the first ‘textbook’ designed for students. These thirteen books on mathematics, and primarily geometry, however, do not at any point define or even utilize the term ‘space’. The most obvious reason for this omission is that the ancient Greeks as a whole did not have a term exactly equivalent to that of space. Instead, the Greeks used either topos, meaning ‘place’, or diastēma, meaning, for the most part, ‘distance’, chōra, meaning ‘locus’, schēma, meaning ‘figure’, and atopos meaning ‘void’ or more literally ‘no-place’. Of course, it is not necessarily the case that one must utilize an actual definition of space in order to present a viable account of geometric objects, yet it seems we would still be inclined, at least in modern thinking, to consider or else just assume geometrical objects to be spatial objects of some sort. Euclid’s response is in no way mysterious; all we have to do is revisit his definitions of the different kinds of objects that he identifies in order to see the manner in which he characterizes them without utilizing the concept of space, at least not explicitly. Instead, perhaps the broadest or most general summation of Euclid’s definitions of geometrical objects is that he relies on a principle of relation, whereby each object involves a particular kind of relation between its constituent parts, particularly, in this case, those of magnitudes and angles. Euclid, however, also does not define or really utilize a term equivalent to ‘relation’ either, but rather more specifically utilizes and defines the concept of logos, i.e., ‘ratio’, yet this is not found in his definitions of such objects themselves. The terms Euclid does mention are directional, e.g., lines that approach one another or that tend away from one another according to angle, as well as, of course, those that are parallel. One point worth noting here is that in Modern Greek chōra can indeed be translated as ‘space’, but again in Ancient Greek the term is usually translated as a ‘locus’ or spot in which an object is located. Euclid first employs another term, namely schēma or ‘figure’ in Book I Def. 14: “A figure is that which is contained by any boundary or boundaries.” We can see how none of these terms can be the

1This paper was presented at the 4th International Philosophy Conference in 2009.
equivalent of a general notion of space as such, for in the case of schēma (which is perhaps the closest term found in Euclid) it is used here to refer to an object, so that the closest conception we could stretch our translation to utilize would still be that of ‘a space’, rather than space itself. Thus, ‘containment’ and ‘boundaries’ pertain here more to the concept of a figure rather than an explicit concept of space. In fact, we can now see how the very notion of ‘Euclidean space’ is itself called into question, since there can only be an inferred statement about what Euclidean space might be since Euclid himself never employs the concept of space in an explicit manner. Thus, while it is not strictly the case that Euclidean geometry is a geometry without space in a strong sense of that phrase, it is nevertheless true that it is a geometry without a definite concept of space, and furthermore one that employs other notions, namely ‘figure’, ‘distance’, and ‘place’ that only imply a general spatial concept or perception that is never explicitly defined or theorized about. It is therefore possible to conceive of a successful geometric system without any explicit notion of space as such.

When we consider this issue more generally, it is not all that surprising to have a geometric system which lacks a definition or theoretization of the nature of space itself. The abstract space of early modern science, particularly that of Newton and any such notion stemming from the Cartesian concept of extension, is far less necessary than we might think. From a philosophical point of view we can think of Russell’s early work *An Essay on the Foundations of Geometry* as an example of what it means to derive epistemological conclusions regarding the status of space itself from modern geometrical systems. The closest ancient Greek parallel to such an investigation is probably that of Aristotle in the *Physics* and the *Metaphysics*, yet even there we do not find an inquiry into space as such, but rather a discussion of *topos*, locomotion, potentiality and actuality. The major difference between an inquiry such as Russell’s and seemingly similar inquiries on the part of the Greeks, is that Russell theorizes about the ontological status of space as such, whereas the Greeks focus on the ontological status of geometric objects, including lines, points, angles, planes and the various figures that such objects constitute.

Even ancient discussions of the infinite do not hypostatize infinite space per se, but rather infinite objects such as lines (including the infinite division of lines), or an infinite cosmos, which is admittedly similar, yet still not identical with a reflection on space qua space. An example of this latter type of discussion is found in Cleomedes’ *On the Heavens*, a late ancient introductory source on astronomy, which highlights the question of whether the cosmos has a limit, and then employs the thought experiment of reaching one’s hand out beyond that limit and asking where is my hand at that point? Simplicius in his commentaries on Aristotle raises the same question.

Another important conception in this regard is that attributed to Democritus, which is also revisited by Aristotle, namely the concept of a ‘void’ filled with atoms. This is probably the closest conception to abstract space that we can find in ancient thought, yet it is already a negative conception of ‘emptiness’, or ‘a-topos’, ‘no-place’ or ‘placelessness’. It is significant also
that Aristotle rejects the possibility of the ‘void’, upholding the notion that an empty space would still need some ‘place’ as a reference point and constituent, similar to the ‘locus’ concept of *chōra* (which is also often translated as ‘receptacle’ in Plato’s *Timaeus*). In all of this, then, we can see how the questions raised are focused on objects or the constituent objects of space, and not on space itself, and thus space itself is simply not in the equation for the ancients. What is dealt with by ancient thinkers, particularly Plato, Aristotle, and Proclus, is the status of geometrical objects. In his *Commentary on Book I of Euclid’s Elements*, Proclus emphasizes the epistemological aspect of such objects by associating them with *dianoia*, thereby placing them in an intermediate category between perception and thought, or *aesthesis* and *nous*.

This latter type of inquiry brings us back to our main interest, which is the question of how to understand and conceive of a geometrical system that does not employ or rest upon an explicit concept of space as such. In a modern context, we seem compelled to wonder, if geometry does not involve a theoretization of the nature of space itself, then what is being theorized about? The ancients seem to engage this question by investigating the status of geometric objects, and in some sense this is similar to the way in which contemporary theories of the cosmos move back into the realm physics, or specifically astrophysics, when dealing with the ‘structure’ of the universe. In other words, modern geometrical questions about the nature of space have eventually come back to the question of what constitutes space, rather than questioning the ontological or epistemological status of space in itself. Thus, we find that the Newtonian or even Cartesian conception of space that we might attach to geometry is a phase of scientific thought that begins long after the ancients and that comes to an end with twentieth century physics. This is significant in that it frees us from having to conceive of geometry as a science that is ultimately based on the question of ‘what is space’, since that notion is neither necessary for the ancients nor for contemporary science. Moreover, it frees us to gain a better understanding of what the ancients were really focused on in their mathematical and geometrical thinking insofar as we cannot falsely attribute or attach the concept of an abstract space to their inquiries. In this regard, one way we might better approach ancient geometrical thought is to consider it a systematic investigation of types of relation, or basically of *logos*, along with the resultant concepts of ratio, proportion, equality, similarity, and commensurability, which are the basic themes found in the successive books of Euclid’s *Elements*. In conceiving of the Euclidean system in this way, we can also better situate Euclid’s work within the larger context of Greek philosophy, science and logic, which in turn allows us to approach Euclid’s geometry in closer connection with thinkers such as Plato and Aristotle, and thereby find greater consistency between ancient philosophy and ancient mathematics as a whole.

Just to be clear, I am not suggesting that the ancient Greeks were somehow different from us in terms of their physical perception or cognitive make-up. I am not putting forth a Whorffian thesis such as the lack of a linguistic term for
space means that there is no perception of space as such. Rather, I am making a much less groundbreaking point, which is concerned with the nature of ancient mathematics in juxtaposition to modern mathematics. Thus, I am not in any way rejecting a Kantian a-priori notion of space in the minds of the ancient Greeks, because it simply does not concern the matter at hand, which instead is the difference between modern geometry and ancient geometry. The point is perhaps best stated when we consider that the concept of ‘Euclidean space’ is in fact a modern concept. Thus the point in question is nothing extraordinary, it is simply a detail that historians of mathematics, philosophy and science (as well as philologists) should consider so as not to take for granted the idea that geometry is always concerned with defining the nature of space, for, as I have attempted to show, this simply was not the case for the ancients. Therefore, the history of mathematics involves a transition from relation-based theories regarding geometrical objects to early modern and modern theoretizations of space itself through geometry. Space as such can, generally speaking, be attributed to thinkers like Descartes, Galileo, Newton and Leibniz, thereby giving rise to the modern geometrical tradition and epistemological considerations such as Russell’s which rely on Kant and his early modern predecessors.

It might be argued that even though the ancients themselves did not speak of ‘Euclidean space’ per se, they nevertheless were working on this notion, and simply didn’t realize it, i.e., that to do geometry at all is necessarily to deal with space qua space, regardless of what one thinks one is really doing. Such a position, however, makes an assertion about geometry itself, and even if it were true, it would not change the fact that the history of geometry involves an ancient approach that was not trying to define the nature of space in the manner of modern geometry. In this brief analysis, what is important, then, is neither the cognitive or epistemological underpinnings of an a-priori conception of space, nor the possible necessity of space in geometry, but rather the particular approach of the ancients toward geometrical objects in terms of relations that did not utilize and did not involve a term or theory of the nature of space itself. In this sense, the ancients did indeed work on a ‘geometry without space’. The contemporary consequences of this suggestion are not concerned with a meta conception of geometry as such, but rather with our characterization of the historical development of geometry and a greater understanding of the ancient Greek approach to geometrical objects which can be juxtaposed to a modern theorizing of space itself through the geometrical tradition. In the end, it is a small point that is modest in scope, yet still important for our continued discussions and sustained scholarship on an accurate account of the history of geometry and science.
A Comparative Study of the Role of Harmony in the Thoughts of Confucius and Plato on the Nurturing of Culture through Music

Mei-Yen Lee

The concept of harmony played a very important role in ancient Chinese and Greek philosophy. Both Confucius (551-479 B.C.) and Plato (427-347 B.C.) regarded harmony as the core concept underlying the process of nurturing cultural mores through music. Over the past few decades, a number of studies on Confucius and Plato have focused on their musical aesthetics and education. However, very few attempts have been made to compare the thoughts of these men as they pertain to the nurturing of culture through music. Thus far, studies on this topic have been somewhat superficial. The fact is that Confucius and Plato had very different views concerning the theoretical basis of this issue. Based on research of their original texts (Confucius’ Analects and Plato’s the Republic, Timaeus, Laws etc.), this paper discusses the similarities and differences between the viewpoints of these two men as they pertain to the concept of harmony and the power of music to nurture change.

The Significance of Harmony in Confucian thought on the Nurturing of Culture through Music

The Confucian tradition began during the Zhou Dynasty (1134 B.C. - 771 B.C.), when “the nurturing of culture through music” was a means to instill a sense of morality in people, to promote noble character traits and integrity, and transform social customs and habits. Morality represents the dominant issue behind the nurturing of culture through music. In the views of the ancient Chinese, morality and music played complementary roles in the development of character; however, morality outweighed music. This cultural ideal gradually became a comprehensive theory during the Pre-Ch’in period, developing into the guiding principles of Chinese culture.

In the Pre-Ch’in period, music and ritual complemented each other in many ways. According to Copleston, Frederick Charles’ study, in the Republic Book I, Plato was influenced by Socrates. However, Books II-X exhibited Plato’s own thoughts. As for Law, it was the writings of his old age.
ceremonial occasions, particularly in sacrificial offerings to the Gods or ancestors in temples, and during weddings, and funerals. In all of the above-mentioned occasions, ritual and music were meant to help people to become more socialized, and develop a higher sense of morality. Music that is solemn and peaceful combined with noble, dignified rituals enables the elevation of one’s spirits. Ancient sages composed ceremonial music, drawing on the profundity of poems, the rhythm of music, and the movement of dance to strengthen moral sensitivity through a sublime synthesis.

Around the end of West Zhou Dynasty, approximately 771 B.C., due to political, social, and ethical collapse, the ceremonies combining ritual and music broke down. The common belief in the ability to nurture culture through music lost its firm foundation and the support provided by the will of the people. This resulted in ceremonial music degenerating into mere formality. As the prominence of ceremonial music gradually waned, popular music began to flourish, signifying a collapse of formalism in music, and a clear indication that public morality was no longer what it had once been. The system in which culture was nurtured through music disintegrated; a process that appears to have been inevitable under those particular circumstances.

Many philosophers followed the tide of fashion during the Pre-Ch’in period, introducing new ideas to prevent the disintegration of the world they knew. Confucius proposed “benevolence” (akin to Buddhist “compassion”) as the guiding principle with which to save the system of ritual and music; however the disintegration continued, due to a failure of the people to value benevolence. Confucius stated:

‘If a man be without the virtues proper to humanity, what has he to do with the rites of propriety? If a man be without the virtues proper to humanity, what has he to do with music?’

How can an individual develop benevolence? Confucius taught his disciples to control their desires through rational conscious thought and establish lifestyles of moderation and harmony by the power of their own will. Confucius regarded the harmony of music as a symbol representing the highest achievement of morality. He felt that the function of cultural development through music is to unite the human mind and creativity into a harmonious union.

Confucian doctrines of cultural development through music involved intrinsic voluntary illumination rather than extrinsic transformation through compulsion. For Confucius, “musical education as a form of moral education” was not simply an idealistic theory; it had a practical application in ceremonial music. Ancient Chinese ceremonial music was originally produced to address political, social, and ethical concerns. However, by the end of the West Zhou period, the collapse of ritual and music resulted in changes to its original

---

content and form.

Ceremonial music was unable to adapt to the new cultural circumstances. New popular music was produced to provide entertainment and cater to the psychological needs of the listener, taking the place of the earlier formal music, which addressed deeper philosophical and moral issues. So-called popular music was able to express people’s feelings and lives in more vivid and varied musical style. In addition, the royal court overvalued the entertainment of popular music accompanied by teams of female dancers, and popular music gradually drifted into “lewd song”.

The transformation of contemporary cultural circumstances resulting in “lewd song” was in sharp contrast to the ceremonial music in the history of Pre-Ch’in musical thought. The collapse of faith in nurturing cultural mores through music was inevitable; however, this did not mean that the cultural essence behind this goal had disappeared. Confucius recognized and greatly appreciated the value of culture during the Zhou Dynasty. Confucius stated, ‘how complete and elegant are its regulations! I follow Zhou.’

Confucius even considered ceremonial music “Great Wu Dance”, which eulogized the virtues and achievements of Emperor Wu in dance accompanied by music, as beautiful rather than good, representing an aesthetic judgment. The value of ceremonial “Shao Music” was more than an art form, it demonstrated moralizing potential. Confucius stated, ‘(Shao Music) was perfectly beautiful and also perfectly good; (Wu Dance) was perfectly beautiful but not perfectly good.

Clearly, the concept of goodness was highly valued in Confucian thinking, echoing similar concepts in Plato’s The Republic, where art had to be sacrificed to a moral imperative. Confucius felt that the value of morality outweighed the appreciation of beauty. Confucian ideals were at the core of the concept of nurturing culture through music, which moved through the ages in a continuous line, leaving an indelible influence on Chinese culture forever.

Most importantly, Confucius felt that cultural mores could be nurtured through music, believing that moral goodness is inherent in the lives of every man. In particular, the value of uniting beauty and goodness through music was regarded as an important element in developing a character of high moral integrity. Likewise, the linking of beauty and goodness was also part of western aesthetics in ancient Greece.

Confucius valued ceremonial music for its ability to cultivate moral sensibility, expanding it beyond that of a simple form of art. In other words, ceremonial music had a practical application with regard to political, social, and ethical considerations. The lyrics of ceremonial music were written prior to the melodies of the hymns. Because the hymns were composed to venerate the

---

merits and virtues of ancestors, the melodies are simple and the rhythms are clear. The hymns are played slowly and in unison. In response to such musical harmony, people become solemn, gentle, and sincere. The hymns of Confucian ceremonial music embody the Confucian belief that moral music is the most beautiful music.

Studies explaining the differences between musical forms of ceremonial music and those of “lewd music” have been insufficient. In referring to the hymns of Confucian music, it is clear that ceremonial music could nurture feelings of modesty and peace as opposed to excitement or depression; in which musical harmony engenders psychological harmony. The final goal of ceremonial music was to nurture beauty and goodness in the soul of the listener. In contrast, non-ceremonial music was regarded as “lewd music,” because it was based on simple pleasures, the corporeal rather than the sublime. Confucius stated in strong terms:

‘Banish the songs of Zheng (“lewd music”)…The songs of Zheng (“lewd music”) are licentious…”¹
‘I hate the way in which the songs of Zheng (“lewd music”) confound the music of the Ya (“ceremonial music”).”²

The other representative figures of Confucianism, Mencius and Xuncius, were both of the opinion that the expression of feelings through music is inherent to our nature. They both embraced the value of music with regard to humanity and Mencius suggested that the emperor compare his feelings with those of others and place himself in the position of the people to bring contentment through the delight of music. The advice offered by Mencius was intended to simultaneously satisfy the musical needs of both the emperor and the people. Mencius believed that the emperor had to empathize (the root of benevolence) with the feelings of others to satisfy their needs, rather than indulge in music selfishly.

In his essay Music Theory, Xuncius indicated that ancient Chinese sages composed ceremonial music to instruct people. Ceremonial music could touch the heart, enabling the listener to become more benevolent and less absorbed in selfish concerns. Xuncius confirmed the significance and worth of nurturing culture through music. A musical education causes the perceptions of the listener to become sharper and more distinct and the movement of the blood and physical energies to become harmonious and peaceful, leading inevitably to a transformation of social customs and habits. In this manner, the world attains a state of universal peace.

The Confucian classic Yue-Ji later indicated that, ‘Music was intended to illustrate virtue,’ ‘music is the blossoming of virtue,’ ‘virtuous airs constituted what we call Music,’ ‘when the music is ended, (due) honor has been paid to

The role of harmony in the thoughts of Confucius and Plato on the nurturing of culture through music

1 These statements regarding music and morality established the foundation of nurturing culture through music in Confucian society. According to Yue-Ji, ‘Music is (an echo of) harmony in the universe; and ceremonies reflect the orderly distinctions (in the operation) of the universe.’ From harmony, all things receive their being; to those orderly distinctions they owe the differences between them. ‘Music has its origin in heaven; ceremonies take their form from the appearances of earth. If the imitation of those appearances were carried to excess, confusion (of ceremonies) would appear; if the framing of music was carried to excess, it would be too vehement.’

As declared by Yue-Ji, sages composed ceremonial music to commune with the harmony of the world. He continued, ‘Music of the grandest style exhibits the same harmony that prevails between heaven and earth; in ceremonies of the grandest form we find the same graduation that exists between heaven and earth.’ In other words, the sages composed ceremonial music in a moral frame of mind, which was meant to elevate the values of musical harmony, uniting at the cosmic level.

Essentially, music in Confucian thought functioned to nurture cultural development. The core of this belief lies in harmony and morality, in which harmony is meant to beautify and perfect the mind (morality) of the listener, and this notion directed the development of Chinese moral philosophy through the centuries. In other words, harmony had two significant aspects with regard to the nurturing of culture through music: one was the characteristics of ceremonial music, and the other was the symbol of a moral life. Through ceremonial music, one’s moral nature is cultivated and elevated, and cultivated individuals expand their moral concerns to encompass social and natural harmony, integrating them into a harmonic union encompassing the entire universe.

The Significance of Harmony in Platonic Thought Related to the Nurturing of Culture through Music

Plato’s thoughts on the nurturing of culture through music are found mainly in The Republic and Laws. For Plato, musical education was the sovereign means to encourage people to embrace the moral good. In The Republic Book III, Socrates stated:

1 Ibid. Available at http://chinese.dsturgeon.net/text.pl?node=10113&if=en [8 February 2011].
2 My translation followed “Chinese Text Project.” Available at http://chinese.dsturgeon.net/text .pl?node=10113&if=en [8 February 2011]. Some of the translation where I didn’t follow the “Chinese Text Project,” I checked the original texts and revised them accordingly.
3 Ibid.
‘Musical training is a more potent instrument than any other, because rhythm and harmony find their way into the inward places of the soul, on which they mightily fasten, imparting grace, and making the soul of him who is rightly educated graceful, or of him who is ill-educated ungraceful.’\(^1\)

However, what kind of music conforms to the platonic ideal of a musical education? How is it possible to attain the goal of a moral education by nurturing culture through music? What musical modes could conform to the platonic ideal of beauty?

In *The Republic* Book III, Socrates stated that “melos” [melodic music] comprises three elements, text, harmony, and rhythm,\(^2\) and only rhythm and harmony could touch the inmost soul and hold it strongly. Socrates emphasized that craftsmen with natural gifts have to follow the ideal of beauty and grace for fear that the souls of unwary young men would be influenced by evil. A musical education can help young men to distinguish beauty from ugliness, and by extension, good from bad. This enables young men to appreciate and delight in beautiful things through musical education. When young men are able to absorb beautiful things into their spiritual lives, their disposition becomes refined and good.

Clearly, the basic concept behind the Platonic belief in nurturing culture through music corresponds to that of Confucian ideals. Both Confucius and Plato emphasized that music is a means to attain spiritual enlightenment, the transformation of social traditions, and political ideals. In other words, the development of culture through music is regarded as a form of moral cultivation, unconfined to the pure aesthetics of artistic beauty. The most important concept, harmony, contains the ideal of moralization within the realm of the entire universe. This notion is the same in both Confucian and Platonic thought.

We found three other views of Plato regarding the nurturing of culture through music that correspond to those of Confucius:

1. Both Confucius and Plato opposed elaborate musical rhythm and varied musical ensembles, for those qualities encouraged people to indulge in physical pleasure. In *The Republic* Book III, Socrates argued that only pure and simple music could cause the soul to react to the ideal expressed by true beauty, and only a soul with a beautiful disposition can create a beautiful musical form. Plato rejected melodies containing dirges and lamentations, linking them to drunkenness, softness, and sloth. He permitted only two types of modes: the Dorian and Phrygian because they ‘most beautifully imitate the utterances of

---

\(^1\)The translation follows *The Republic* by Plato Translated by Benjamin Jowett, Late Regius Professor of Greek in the University of Oxford. Available at http://ebooks.adelaide.edu.au/p/plato/p71r/ [8 February 2011].

men failing or succeeding, the temperate, the brave.\(^1\) Plato also believed that polychordia or panharmonia modes were unnecessary. As for musical instruments, he preferred the Apollo (7 stringed lyre) to that of the Marsyas (double flute).\(^2\)

2. Both Confucius and Plato emphasized that rhythms have to be confined in a particular manner. In *The Republic* Book III, Socrates stated:

\[\text{‘We must not pursue complexity nor great variety in the basic movements.’}\]

\[\text{‘……the rhythm and harmonia follow the text and not the text these. (The text) follow and conform to the disposition of the soul.’}\]

\[\text{‘……good speech, then, good accord, and good grace, and good rhythm wait upon a good disposition……. the truly good and fair disposition of the ethos and the mind.’}\]

In other words, the aim of musical education is to cultivate a truly good ethos and fair disposition in the mind. This view is the same as that of Confucius. Both Plato and Confucius ascribed the judgment of beauty to the judgment of morality.

3. In *Laws*, Plato used dialogue between Athenian Stranger and Cleinias for his discussion on the educational value of music. Book II, the Athenian Stranger says:

\[\text{‘…the excellence of music is to be measured by pleasure. But the pleasure must not be that of chance persons; the fairest music is that which delights the best and best educated, and especially that which delights the one man who is pre-eminent in virtue and education. And therefore the judges must be men of character, for they will require both wisdom and courage.’}\]

Confucius believed that the nurturing of culture through music instructs a man to act morally, as a gentleman. Only a cultivated man can realize and appreciate “ceremonial music.”

Certainly, there are some differences between the Platonic and Confucian

---


\(^2\)Lippman, E. A. (1964). *Musical Thought in Ancient Greece*. New York: Columbia University Press, p19. According to Lippman, Edward A., In support of the thesis that the name Apollo means “moving together,” Socrates points to the fact that Apollo is the god of harmony; musicians and astronomers both declare that he makes all things move together by a harmonious power, whether in the harmony or concord of song or in the poles of heaven.


\(^4\)Ibid. p13.

\(^5\)Ibid. p13.

viewpoints on the nurturing of culture through music:

1. In *The Republic* Book III, Socrates stated that:

‘And further, because omissions and the failure of beauty in things badly made or grown would be most quickly perceived by one who was properly educated in music, and so, feeling distaste rightly,...’¹

‘...it has certainly made a fitting end, for surely the end and consummation of all musical things is the love of the beautiful.’²

Confucian ceremonial music was only performed for the nurturing of culture through music rather than for the appreciation of music as art. Confucius did not discuss how musical knowledge influences the recognition of artistic works and nature.

2. In *The Republic* Book III, Socrates stated that education in music and gymnastics are ‘both chiefly for the sake of the soul.’³ The soul of the man thus harmonious is sober and brave.”⁴ Plato emphasized that music and gymnastics complement each other to establish the base on which to develop a perfect character. In *The Republic* Book III Socrates stated that ’simplicity in music begets sobriety in the soul, and in gymnastic training it begets health in bodies.’⁵ Confucius did not mention music and gymnastics in the same breath. Confucius simply listed archery as one of six kinds of art education including: rituals, music, archery, chariot, calligraphy, and math.

3. In *The Republic* Book IV, Socrates emphasized that music and gymnastics, preserved in their original form, should not be innovated, for any innovation would result in danger to the entire State, and ought to be prohibited. Similar viewpoints are found in *Laws* Book III, where the Athenian Stranger says:

‘...as time went on, the poets themselves introduced the reign of vulgar and lawless innovation. They were men of genius, but they had no perception of what is just and lawful in music; raging like Bacchanals and possessed with inordinate delights-mingling lamentations with hymns, and paeans with dithyrambs; imitating the sounds of the flute on the lyre, and making one general confusion; ignorantly affirming that music has no truth, and, whether good or bad, can only be judged of rightly by the pleasure of the hearer.’⁶

²Ibid.p16.
³Ibid.p17.
⁴Ibid.p18.
‘...in music there first arose the universal conceit of omniscience and general lawlessness; freedom came following afterwards.’\textsuperscript{1}

From the viewpoint of Plato, music should not be allowed to change for fear that the political stability of the state might be undermined. This viewpoint was somewhat different from that of Confucius. Although Confucius preferred ceremonial music and found lewd music distasteful, his emphasis was on the character rather than the form.

Harmony was commonly esteemed by Confucius and Plato with regard to the nurturing of culture through music. Plato believed that harmony is the incarnation of the human soul in cosmic order, representing the ideal of beauty and goodness. In \textit{Timaeus},\textsuperscript{2} there is an important paragraph:

‘The body of the heaven has been created visible; but she (harmonia) is invisible, and as a soul having part in reason and harmonia, is the best of things brought into being by the most excellent of things intelligible and eternal.’\textsuperscript{3}

The soul is derived from reason and harmony, as is the ideal of music. The ideal of harmony is the mutual source of music and soul in the sense of divine inspiration. Therefore, eternal harmony may exist in the human soul if it is cultivated through musical education.

From the viewpoint of Plato, music embodies the spirit of harmony and cultivates the soul of individuals. A real musician is one capable of perfecting his soul and the souls of others to conform with the ideal of harmony. Plato felt that the ideal of harmony is the foundation of the human soul, existing in an abstract metaphysical realm rather than the phenomenal world. It is absolute, far beyond the audible music performed by well-trained musicians. In the Confucian view, harmony could be reached in one’s inner world through introspection, representing the first step toward attaining wisdom. By continually exerting one’s self to cultivate a moral character and concern for others, the entire universe remains in a perfect harmonious union.

\textbf{Conclusion}

In conclusion, both Confucius and Plato regarded the role of music as a means to nurture cultural development, through the attainment of harmony. From the perspective of both men, harmony was virtually synonymous with

\textsuperscript{1}Ibid.
\textsuperscript{3}Ibid. p22.
morality, and viewed as a means to beautify and perfect the human mind as it strengthens one’s moral sense. However, Plato and Confucius followed different paths to attain their goals.

Harmony, in the Confucian ideal of nurturing culture through music, enables people to achieve the utmost meaning from life. The final goal for Confucius was concern for others and the perfect union of the entire universe. Plato felt that music is an imitation of harmony, and that people can be developed morally through musical education and a love of beauty. Plato’s view is very different from that of Confucius in that he considered music a reflection of the harmony of the universe. In ancient Greek cosmogony, the universe was a creation based on the transformation or evolution of chaos into harmony, the divine universal principle.

Confucius believed that harmony was derived from the moral practice of the human mind and affirmed that moral goodness is inherent in the human soul. The platonic ideal was derived from abstract thought, rather than lifestyle. This is the most obvious difference between the doctrines of Plato and Confucius with regard to the nurturing of culture through music. Finally, the author argues that the differences between the concepts espoused by Plato and Confucius regarding harmony in the nurturing of cultivation through music were metaphysical in nature, based on divergent views regarding the nature of man and the cosmos.

References

A Comparative Study of the Role of Harmony in the Thoughts of Confucius and Plato on the Nurturing of Culture through Music


Any discussion in regard to the essence of philosophy would be incomplete if it did not consider the meaning of the Ancient Greek word \( \phi i l o\sigma o\phi i a \). The meaning of the word \( \phi i l o\sigma o\phi i a \) is somewhat cloudy, particularly since its component, the word \( \sigma o\phi i a \), has not been clearly defined by scholars. Our research, however, concludes that the word \( \phi i l o\sigma o\phi i a \) surmises the problematical character of wisdom as the means of cosmogenesis cognition in principle. The problematic character can be formulated in the following question: Was the cosmos self-generated or created?

This world, which is the same for all, no one of gods or men has made; but it was ever, is now and ever shall be an ever-living. Fire, with measures kindling and measures going out.

Heraklit (Frg. 30. John Burnet’s translation)

Let me tell you then why the creator made this world of generation.

Plato (Tim., 29d, 30a. Benjamin Jowett’s translation)

It is generally known that the word ‘philosophy’, which comes from Greek, literally means ‘love of wisdom’. This translation is usually taken as the absolutely clear, comprehensive sense of the word. Such a perception of the word ‘philosophy’ may be explained, at least partially, by the fact that the gnosiological plane of content, which dominates in the common translation, is in tune with the cognitive function of the philosophy or methodology of science. This function determines the social prestige of philosophy in the current age, but it also determines the danger of reducing philosophy to the form of scientific cognition as well as the danger that its specific identity may be lost. Philosophy can be as solid as the philosophy of science or the methodology of science only when it is solid as philosophy proper. Perhaps that is the reason why, in the modern age of the twentieth and twenty-first centuries, there is a tendency to raise a question best described as thorny, in the manner of Plato: What is philosophy per se?

The statement of a question of essential philosophy leads us back to antique philosophy as an archetype; as an ever-living, effective pattern of philosophizing. In the reconstruction of an archetypical mode of philosophy, an essential part should be assigned to elucidation of the meaning of the ancient Greek word \( \phi i l o\sigma o\phi i a \).

It was M. Heidegger who obliged the modern day, when he indicated that
the way to answer this question should begin with elucidation of this word meaning. However, this approach, as Heidegger fairly notes in his article titled *What Is Philosophy?* (Heidegger, 2001), is being ignored because the word ‘philosophy’ – a calque of an Ancient Greek word – appears to be worn out. In order to make it ‘speak Greek’, one must deal with its Ancient Greek original. Heidegger probably means that the implicit, lying under the surface of mere translation, meaning of the word *φιλοσοφία* can be revealed only in the context of Ancient Greek culture. His analysis shows that the deeper meaning of the word *φιλοσοφία* lies in the field of ontology, and suggests the problem of distinction between categories of being and things in existence. This shift of meaning from the gnosiological to the ontological is important, as Heidegger correctly observes.

Heidegger’s reconstruction gives too abstract a result, in any case. This is probably because he limits his research area to the philosophical texts of the Ancient Greeks. Of course, the word *φιλοσοφία* for antiquity is a neologism, since it serves to distinguish philosophy from the previous styles of thinking – sophistication and mythologizing; it serves to self-affirm a new, philosophical type of thinking about the world. The effect of such a novel meaning is achieved by tying together two words – *φιλία* and *σοφία* – both of which are deeply rooted in the pre-philosophic cultural tradition. Thus the word *φιλοσοφία*, being a neologism, also includes specifically articulated meanings of the corresponding words, such meanings referring to primordial socio-cultural and mythological ancestry. The philosophers of Ancient Greece, as direct heirs of pre-philosophic cultural tradition, supposed the pre-philosophic layer of the word meaning *φιλοσοφία* as taken for granted, self-evident and therefore requiring no special discourse. However, in the course of time this self-evidence was lost and now special research efforts are necessary in order to discern what is behind the ‘worn out’ meaning of the word *philosophy*, or *love of (aspiration for) wisdom*.

**Φιλία and Σοφία:**
*Word Concepts in the Mythological Context of Word Names*

As we consider the question of the meaning of the word *φιλοσοφία*, it is necessary to note that it is constructed not from the name words, denoting mythological divine persons but from word concepts. Why was a word consisting of word concepts, notably *φιλία* and *σοφία*, invented to denote a new type of cognition instead of taking some grammatical form such as a combination of the words *Ερως* (embodiment of passionate love) and *Ἀθηνᾶ* (one of the embodiments of wisdom)? Obviously, it is because the new type of thinking was initially, mainly constituted as conceptual, unlike mythological thinking, which was mainly figuratively personifying.

However, the fact that Ancient Greek philosophers preferred word concepts not only for cognitive causes but even to denote their type of thinking does not mean they could ignore the meanings of mythological personification images,
particularly those that presented the meaning, content and context of the words φιλία and σοφία, which later formed the word φιλοσοφία, within the framework of mythological consciousness. It was impossible to change the previous tradition by breaking away from it. That could only be done by transforming the tradition. In our case it means that the act of combining the word concepts φιλία and σοφία into one word concept φιλοσοφία was a means of demythologizing the context, and that it was dependent on the context meanings of the source words. Moreover, the initial meanings of the word concepts φιλία and σοφία had to be presumed by the philosophical consciousness. That is why, if we want to reconstruct the meaning of the word φιλοσοφία, we should first reconstruct the initial meanings of the words from which it is composed. Now let us consider the etymology of the word φιλία.

Φιλία

E. Benveniste (1969) pointed out that, for the word φιλία (philía) “friendship, love,” and for a number of paronyms: φιλότης (philótēs) – having the same meaning as φιλία; φίλος (philós) ‘friend’; φιλεῖν (phileîn) ‘to love, to be friends with, to kiss’, etc., the source word is the adjective φιλός ‘loved, dear’, which originated in pre-Homeric times. Initially the word functioned within the limits of a group, based on kinship or a family collective. In that context the word φιλός meant personal, friendly feelings of friendship or love, which a person as a member of a certain social collective would feel toward some of his kinsmen in the circle of direct communication. Loving in the sensually affective (being characteristic of sexual attraction) sense appears in the general meaning of this word only as a particular point, in connection with addressing φιλός to a spouse; only as a particular point of the general meaning of friendship.

In Homeric times φιλός relations transcended the limits of a group, based on kinship and affinity. In this process there occurred the formation of a cluster of φιλός paronyms and the universalization of their meaning on the side of application to increasingly diverse uses that were distant from direct social interests and obligation objects (not only people). Benveniste fully illustrates this process by way of many examples from Homer’s Iliad.

Based on the examples Benveniste gives, one can see another aspect of the meaning of the words, derivative of φιλός, but it is one that Benveniste does not actually note. The matter is that as far as the universalization of these word usages occurred, in addition to the fact that some time earlier ‘friendship’ had begun to dominate over ‘love’ in φιλός, the nuance of lust in ‘love’ became further sublimated. The motives of feelings, expressed by these words, were thus idealized. For example, ‘love’ became love for one’s native country, which can be seen from an expression in Homer’s Odyssey: ‘φίλημα καὶ πάτρια γαῖαν’ (‘in the loved (dear) native land’) (Odys. XV, 65). It means that the word φιλία (φιλότης) in particular, which is the most emotional and expressive among the words, being derivative from φιλός, presupposes attraction, the impulse to
which comes from the importance of one or another object to an individual. In other words, φιλία is an unselfish (not physiological) attraction to an object, where the object serves for an individual as what Plato called φιλόθεν (common good): the indissoluble unity of truth, goodness and beauty.

It is necessary to note that the logic of universalization of the meaning of φιλία (φιλόθης) lends extremely broad scale to the strength and the sphere of action of ‘love’. There is at least one significant example of attaching cosmic scale to love – φιλότης – which is Empedocles’ doctrine. This example is from the fifth century B.C., i.e., the formative period of philosophy, but clearly the tendency to use the word φιλία (φιλότης) in such a way would have appeared much earlier. Empedocles conveys the essence of his cosmogonic theory in the poem On Nature (περὶ φυσικῶς) in the following thesis (my translation – V. M.): ‘Now under the influence of Love (Φιλότητι) they unite in the single cosmos, || now they fly asunder under ferocious Strife (φοροῦμενα Νείκεος), || until they, adnatated before in the integral Universe (τῷ πᾶν), fall apart <…> . But as far as this continuous alternation ends not, || so far as they always exist, for they go by immovable circle’ (Empedokl. Frg. B, 26, 5 - 7; 11 - 12).

Although Empedocles tears φιλότης – φιλία away from man as a subject of this love, it in turn destroys the initial inner sense of this word on the side of wholeness, necessary fullness and depth. In this respect it turns out that Empedocles’ word usage did not correspond to the Greeks’ general understanding of the meaning of the word φιλότης – φιλία, for if φιλία was not human love in the general understanding there would have been no general understanding of philosophy as a distinctly human affair, being human loving aspiration for wisdom.

Generally, Empedocles’ utmost broadening of φιλότης – φιλία meaning to the scale of the universally cosmic Love, bringing the pre-philosophic tendency of universalization of this meaning to the completion of natural logic, also adequately serves the cause of philosophical self-determination. For only such scale of love-inclination is commensurate to sophia (wisdom). Σοφία itself, as we will see, has a universally cosmic dimension.

Σοφία as opposed to τέχνη

The etymology of the word σοφία (sophia) is difficult to reconstruct, despite the fact that many research publications are devoted to this subject. This is because once getting caught in the thick of philosophical thought, and in turn theological thought, this word in the course of cultural-historic process appears to be overloaded with various semantic strata, as a result of which the word’s original meaning is almost hopelessly obscured.

A valuable contribution to the research of etymology of the word we are interested in was made by V.N. Toporov.1 We will rely on his final work on that

---

1Vladimir Nikolayevitch Toporov (1928 - 2005) was a Russian linguist, semioleologist, mythologist and philosopher. He was an academician of the Russian Academy of Sciences, a
subject (Toporov, 1995).

In the Ancient Greek mythopoetic works, the word σοφία often has a meaning coincident with that of the word τέχνη (téchne) (art). Τέχνη (art) was conceived by the Greeks as such an aptitude, due to which a thing (a work of art) could be made skillfully, in a workmanlike manner. It was conceived as an aptitude, expressed in perfect skills, sophisticated abilities, i.e., as an aptitude we could call the aptitude to act in a technically, technologically perfect way. That is why the ancient Greek word techne underlies the words technics, technology and so forth in the European languages but does not underlie the word art. However, the ancient Greeks used the word téchne for each ‘art’, whether for poetry or, for instance, architecture. In the meaning, synonymous with τέχνη, σοφία is consequently an aptitude to create technically perfect things. But the meanings of the words σοφία and τέχνη are of course not only synonymous but also different.

As V.N. Toporov pointed out, despite the seemingly obvious synonymy of the words σοφία and τέχνη in the mythopoetic works, one could discern a subtle but significant difference in their meanings. The word τέχνη serves the designation of the efficiency, skillfulness and craftsmanship of a person as an individual who creates certain remarkable, beautiful and at the same time useful things. ‘Τέχνη’ is acquired and passed on by means of teaching. In the ancient texts a master, a creator-demiurge (‘demiurge’ in the culture of Ancient Greece was any craftsman: a carpenter, a shoemaker, a potter, etc., as well as a talented poet or musician, etc.) is called wise, but his handicraft creativity (τέχνη) is called wisdom in the meaning of σοφία only when he has taken his art from the gods. For example, in one of the texts it is said (the translation is mine – V. M.): ‘From the gods comes every virtue of the mortals: wisdom, and strength, and eloquence . . . .’ (Ἐκ θεῶν γὰρ μαχαναὶ πασαὶ βροτάεις ἄρεταις, καὶ σοφοὶ καὶ χερσὶ βιαται περίγλοσσαι τ’ ἐξεφι . . . .) (Pithia, I, 41 - 2).

Thus the meaning is that σοφία (wisdom) is τέχνη (art) of divine origin. That is so, particularly when we consider how sophia (wisdom) is presented in mythopoetic works. For such a philosopher as, for example, Heraclitus, in whose works we find statements about σοφία, ‘Wisdom is to speak truth and consciously to act according to nature’ (translated by John Burnet) (Heraklit. Frg. 112). Heraclitus removes the points, formed by mythological divine personification of world natural being, from sophia (wisdom). However, in the mythopoetic tradition and in Heraclitus’ statement σοφία (wisdom) is generally presented as the mind’s aptitude, concerned with ‘the supreme and most universal power, which rules everything in the world’, as V.N. Toporov member of the editorial boards of a number of Russian and international scientific journals (particularly International Journal of Poetics, Kodikas, Proverbium and Arbor Mundi), and a member of international scientific societies (particularly Academia Europaea, the International Semiotic Society, the International Language Origin Society and the International Philosophic Peirce Society). V.N. Toporov reconstructs mythological conceptions of the peoples of the world as an integral universal system of signs and symbols. See: Toporov V.N. (Translated by Rudy S.) On the semiotics of mythological conceptions about mushrooms, in the book: Semiotica. Vol. 53, Issue 4, 1985. pp. 295 - 358.
V.N. Toporov traces the origin of the Ancient Greek word \( \sigma\omicron\varphi\iota\alpha \) from the Indo-European word \( *su\breve{\upsilon}\breve{\omicron}\breve{b}h\breve{\eta} \), meaning both isolation (this sense is particularly reflected in the derivative of \( *su\breve{\upsilon}\breve{\omicron}\breve{b}h\breve{\eta} \) in the Slavic words \( osob \), \( osoba \), etc.) and inclusion in a wider generality, which ‘osob’ of course acquires. Such origin of the word \( \sigma\omicron\varphi\iota\alpha \) clarifies the point of its meaning content, where \( \sigma\omicron\varphi\iota\alpha \) (wisdom) as the mind’s aptitude to be concerned with ‘the supreme and most universal power’ also serves as the mind’s ability to see the overall all-unity, the unity of ‘all in all’ in this ‘power’. This all-unity possesses such fullness that it appears before the mind as self-sufficient and, consequently, a self-generating unity. The ability of a sophia mind to see self-generating universal (i.e., cosmic) unity is equivalent to the ability of self-deepening, self-consciousness and self-knowledge.

**Σοφία in the Dialectical Unity with τέχνη**

Thus we, having more or less discriminated the meanings of the words \( \sigma\omicron\varphi\iota\alpha \) and \( \tau\acute{e}\chi\eta \), must then answer the questions: Why do these words in mythopoetic texts frequently accompany each other, and why are they often used as synonyms? Without answering those questions we cannot unravel the mystery of the meaning of the word \( \sigma\omicron\varphi\iota\alpha \) (sophía).

It is clear from what has already been said about the divine origin of \( \sigma\omicron\varphi\iota\alpha \) (wisdom), according to the mythological conceptions of the Ancient Greeks, that the human mind possesses wisdom if it is concerned with the divine, universally cosmic activity. However, such divine activity is fundamentally important in reference to the activity of a god or a goddess.

The matter is that the difference of gender among gods acts as a primary concept in mythology for many other classifications of positions and roles of gods in the pantheon. That is why we should pay more attention to gods’ gender differences when we put forth questions about the meanings of \( \sigma\omicron\varphi\iota\alpha \) and \( \tau\acute{e}\chi\eta \) as well as questions concerning the reasons for their synonymy in the mythological thinking.

The gender differences of divine mythological characters are comparable to two different types of mythological conceptions of the cosmos origin, i.e., about how the universe became the cosmos, or the orderly universe. One type is the conception of the origin of the universe by means of generation or, to be more precise, the self-generation of the state of cosmos by the initial universal chaos. The other type is the creation of the cosmos from its predecessor, chaos. This is clearly seen when we compare the images of female and male gods, who play significant roles in the pantheon.

In the mythology of Ancient Greece, by the time it was recorded, a special role in the female line of the pantheon was played by Gaia (earth) (typological equivalents to Gaia are the images of the Mother of Gods, Mother Nature, etc.), while in the male line it was played by Zeus. Gaia, in particular, personifies the self-generating, universally cosmic origin. She, if we follow the mythological
On the Issue of the Word Φιλοσοφία in its Original Sense

tradition as recorded by Hesiod (below cited by Theogony (Hesiod. Theogn. 115 - 129) in the translation of Hugh G. Evelyn-White), was born (self-generated) after chaos. (We agree with many commentators that, in this episode, Gaia is the personification of the chaos potential to generate the cosmos.) Then ‘she bore also the fruitless deep with his raging swell, Pontus, without sweet union of love,’ hills, nymphs, and she ‘first bore [for herself – V. M.] starry Heaven’ (Uranus) ‘to cover her on every side’. Contracted with Uranus, she formed the pantheon and everything else in the world.

The mythological idea of the self-generation of the cosmos by the universe is probably equivalent to the feminine ability to give birth: The ability of self-generation is the deified ability to give birth. If σοφία is, as established in the above-stated etymological reconstruction, the ability of the divine mind to see the universally cosmic unity as self-generating, then it is clear that sophia (mind) is the natural (from a mythological standpoint) property of the mother goddess, in the Greek mythology, first of all, of Gaia, the Mother of the Cosmos, and then of the goddesses, being functionally derivative of this character.

The Greek mythology more expressively conveys the idea of the origin of the cosmos by its creation in the image of Zeus than in the image of any other male god. Zeus is, first of all, the demiurge of the cosmos. However, Zeus in the Greek mythological tradition is not necessarily secondary to Gaia in genealogical terms, meaning that he is not necessarily her ‘grandson’. For example, in orphic poems and hymns, as opposed to Homer’s and Hesiod’s works, Zeus is often genealogically primary. The orphic text of the Derveni Papyrus first gives the myth about Zeus’ origin from Gaia – the same as in Homer’s and Heriod’s works, Zeus is often genealogically primary. The orphic text of the Derveni Papyrus first gives the myth about Zeus’ origin from Gaia – the same as in Homer’s and Heriod’s works – but then it is told that (further on the Derveni Papyrus is cited in Gábor Betegh’s translation), Zeus ‘swallowed’ the previous gods (Derveni papyrus. Col. XIII, 5) and it appears that ‘Zeus was born first’ (Derveni papyrus. Col. XVIII, 12 - 13). And the origin of the cosmos occurs in a different way than is described by Homer and Hesiod. Now Zeus himself stands at the origin of the world: ‘Zeus the head, Zeus the middle, and from Zeus all things have their being’ (Derveni papyrus. Col. XVII, 12). ‘Zeus the king; Zeus who rules all <…>’ (Derveni papyrus. Col. XIX, 10). ‘(It has got the name) Harmonia, because many of the beings got fitted to one another’ (Derveni papyrus. Col. XXI, 11 - 13). All this work (ἔξγνλ) of creation of the cosmos, concluded with the creation of the sun, Zeus did due to his all-powerful mind (πάλη Δηοθξῆλ) (Derveni papyrus. Col. XXV, 14).

In the mythopoetic texts one can find examples of the fact that the mythologem of Zeus’ being a ‘grandson’ of Gaia is absent and Zeus at once acts as ‘the first and the last’. As is said in one of the orphic hymns to Zeus (translated by Thomas Taylor): ‘Our pray'rs and expiations, king divine, for all things round thy head exalted shine. || The earth is thine, and mountains swelling high, the sea profound, and all within the sky’ (Orphei himni. XV). Zeus possesses, first of all, a ‘technical’, creatively skillful mind (πέχανη).

One can rightly conclude that, at least, those female goddesses, who are not devoid of the feature of wisdom, realize one or some other aspects of Gaia: the
function of (self-) generation of the cosmos and, respectively, they chiefly possess wisdom in the sense of σοφία. On the other hand, male gods, masters and craftsmen, realize various aspects of Zeus’ function, i.e., the functions of world construction, and respectively they chiefly possess wisdom in the sense of τέχνη.

The described state of things is obscured by the fact that, as the mythology researchers M. Eliade, C. Kerényi, C.G. Jung, Joseph Campbell and others have pointed out, mythological creatures are androgynes, i.e., they combine both sexes. Moreover, it is the cosmogonical gods who generally have an androgynous nature (Jung and Kerényi, 1949). It is also important to keep in mind that those gods, who chiefly act as male and the goddesses who chiefly act as female, prove to be androgynes (Eliade, 1991).

The fact that Gaia is a female androgyne is clear because of her bearing Uranus ‘for herself’. Thus she also contains a male basis within her female nature. Consequently, Gaia is not only mother in the version of the cosmos origin by its generation, but she is the father as well, and consequently she is the creator of the cosmos. That is exactly what is said in one of the orphic hymns (translated by Thomas Taylor): ‘Father of all <…> and mother kind’; she bears the cosmos, but she can also create it, ‘Much-mechanic mother’ (Orphei himni. X, 19; 21).

The fact that Zeus is a male androgyne is particularly clear from the orphic texts as well. Thus, in one of the myths he bears Aphrodite (Orph. fr. 183), and in the other Aphrodite and Eros (Orph. fr. 184). There is a myth in which Zeus generates himself and then the other gods (Orphei himni. VIII). Additionally, there is an idea that he alone bears the cosmos in toto (translated by Thomas Taylor): ‘All-parent, principle and end of all, whose pow’r almighty, shakes this earthly ball’ (Orphei himni. XV, 7).

Everything that was said about the genealogical primacy of the now-female, now-male line of the pantheon, and also about the androgyny of cosmogonical gods, is also clearly seen in the orphic tradition and is generally true, albeit not so evident, for the whole Greek mythology (and probably others). The proof of the preceding is provided in the images of the general Greek gods with the hypertrophied second sex. In the female line it is the image of Athena, and in the male it is Dionysus. Athena is a virgin woman, i.e., the woman who does not realize the divine female function of cosmos generation. She is, on the contrary, so manful that she competes with Zeus and is a goddess of war. She patronizes craftsmen-demiurges, i.e., she takes part in the creation of the cosmos. Dionysus is the most feminine Greek male god, based on his constitution. Dionysus is the ‘dying and resurrecting god’ who personifies spring’s revival of nature, i.e., he participates in the generation of the cosmos, in the manner of feminine gods.

Androgyny and precedents of the second-sex hypertrophy of the cosmogony gods indicate the mutual complementarity and deep mutual penetration of the two versions of the cosmos origin: the spontaneous generation model and/or the creation model. Naturally, this character of relationship also exists between the two kinds of wisdom: σοφία wisdom and
On the Issue of the Word Φιλοσοφία in its Original Sense

τέχνη wisdom. Thus the matter is that σοφία inwardly presupposes τέχνη as its mediation and compensation, and in turn τέχνη contains σοφία as its inner point.

The meaning of the word σοφία therefore has a deeply problematic character: What is primary: σοφία or τέχνη? In the ontological vein, how did the cosmos come into existence? Did it occur spontaneously or was it created? This problem simply did not exist for the mythological conscience. Different versions coexisted without either being considered an alternative. It was philosophy that began to realize these questions as a matter of principle.

Φιλοσοφία and Philosophy

We can now conclude by providing an outline of the meaning of the word φιλοσοφία based on the reconstructed meanings of its constituents φιλία and σοφία.

The meaning of the word φιλοσοφία implies that it is discussed as an intimate (referring to a person, as included in the system of social bonds), value-motivated inclination to such quality of mind activity that, due to its capacity for self-deepening and self-knowledge, it could be concerned with the cosmos origin process. There the mind aptitude is implied, which either includes the art (technique) of cognitive activity or, on the contrary, is included in this art as one of its points, seeing a problem in the cosmos origin: Either it comes from self-generation by the primary universal chaos, or it is created from the chaos.

The philosophers therefore insisted that they were not wise men but only wisdom-lovers; that they had realized the absolute problem. No matter which of two possible answers a man gives, the question will always be problematic, but the search for the answer remains and will forever remain an essential one. In other words, that is the reason the philosophers needed to combine the word φιλία, meaning the essential value, with the word σοφία, into one word: φιλοσοφία.

Of course, it makes all the difference whether the image of philosophy is simply implied by the meaning of the word φιλοσοφία or the image of philosophy is being realized, deliberated and posited within the foundation of a detailed treatment of what philosophy is. The first, theoretically elaborated treatment of what philosophy is belongs to Plato. The theoretical image of philosophy, presented in Plato’s doctrine, not only develops the meaning, as supposed by the word φιλοσοφία, but it also somewhat transforms that meaning. It is therefore necessary to elucidate this word meaning, but that alone is not enough. It is also necessary to understand the theoretical image of philosophy, as it was developed in Ancient Greece. Such a topic, however, is beyond the bounds of this article.
Bibliography


*Orphicorum fragmenta*. Ed. O. Kern. Berolini: Apud Weidmannos. 1863. [In Greek].
CHAPTER TEN

Friendship and the Common Life: Aristotle’s Contribution for a Modern Utilitarian Society

Philip Matthews

At the heart of Aristotle’s *Nichomachean Ethics* (NE) there is an awareness of the role that friendship plays in the formation of moral character. This ontological aspect of moral deliberation stands in stark contrast to the decision-making protocols currently in vogue for most forms of practical ethics, particularly the style of preference utilitarianism adopted in Australia.

As the twentieth century began to unfold, a series of achievements in science caused a *technical anxiety* that had hitherto been unknown. Two major developments serve as book-ends to this twentieth-century period of technical anxiety, Einstein’s theories of special and general relativity (1905, 1916) and the completion of the human genome map (2000). One of the practical applications of $E = mc^2$ was the development and use of atomic weapons.\(^1\) At the other end of the century, advances in the genomic sciences produced a technical anxiety over the hereditary risks associated with transgenic species, the moral issues associated with cloning, and the ‘big brother’ risks associated with genetic information.\(^2\) These technical developments are not easily resolved by the Modern approach to complex moral issues because their remains a conceptual incommensurability between rival moral theories (consequentialists and non-consequentialist).

Faced with this technical anxiety a new discipline evolved within moral philosophy that attempted to side-step the epistemic problems associated with conceptually incommensurable theories. Known variously as applied or practical ethics, this new discipline sought to overcome the standoff between competing theories of ethics by focusing on decision making protocols. Most of these new protocols advocate a composite approach whereby the rational agent takes on the role of an impartial moral agent and incorporates aspects of both consequentialist and non-consequentialist theories of ethics in the decision

---
\(^1\) For several decades the threat of a nuclear exchange between feuding nations produced a social anxiety that psychiatrists labelled *nuclear anxiety*. See Meacham, S. (1964) ‘The Social Aspects of Nuclear Anxiety.’ *American Journal of Psychiatry* 120: 837-841.

\(^2\) The *National Human Genome Research Institute* acknowledged public anxiety over genetic research and devoted US$18 million dollars annually to what they refer to as the ‘Ethical, Legal and Social Issues’ (ELSI) associated with mapping the human genome, this exceeds all other funding previously devoted to ethical, legal, and social issues in health care.
making process. Advocates of this approach argue that practical ethics of this type can solve moral dilemmas in a way that had not been possible using traditional ethical theories. However, the first generation of bioethics produced decision making protocols that are too simplistic for the type of complex moral dilemmas that moral agents face in a pluralist society. The second generation of bioethics will hopefully be more philosophically rigorous.¹

This paper will make two related claims as a contribution to this second generation because bioethics needs to rediscover aspects of the moral life that it has hitherto ignored. First, Aristotle’s concept of friendship is required both for the flourishing of a modern pluralist society and for the formation of moral character. A moral agent is not an automaton and decision making that is removed from a person’s character is necessarily diminished. Second, a flourishing pluralist society engages a moral agent in a type of conversational pragmatism over thin values that we hold in common and thick values that are defined by tradition or practice.

Section 1: Friendship and Civic Responsibility

Something like Aristotle’s concept of friendship is necessary in a modern pluralist society because friendship is a formative relationship that develops moral character and thus enables us to be good citizens. I say something like Aristotle’s concept of friendship because I don’t think his description of the various types of friendship is the final word. In Books VIII and IX of the NE Aristotle provides useful descriptions for three types of friendships: friendship of pleasure, friendship of utility, and friendship of character or excellence.

The aim of this paper is not to unpack what friendship means in any detail but to simply reflect on why Aristotle thinks that friendship, in its various forms, is necessary for the formation of a morally good person. The reason for this is because friendship is an activity that informs other virtues. Friendships of moral character require a level of intimacy not required for friendships of pleasure or utility but Aristotle rightly shows that these close friendships are also quite rare. A moral agent will probably only have a handful of friendships of moral excellence, primarily because these relationships take time to cultivate, probably decades. One of the major benefits of this type of friendship is that it involves a type of intimacy that circumvents the personality masks we wear when engaged in other forms of friendships. However, this level of intimacy is neither possible nor required for the types of relationships that one needs for many daily activities, such as earning a living. Friendships of pleasure and utility are just as pivotal for living well in the polis because they are more commonly required in the types of relationship we share with most others.

Aristotle articulates better than most why the study of ethics is first and

foremost a practical discipline. For him the object of moral enquiry is not knowledge for its own sake but rather for the practical benefits that flow from this type of enquiry, within the life of the polis. In the opening chapter to the NE Aristotle explains why politics is the ‘master art,’ primarily because ‘politics uses the rest of the sciences.’ In Book VI he expands on the intricate relationship between politics and ethics by first showing the difference between scientific knowledge and practical wisdom. For Aristotle, scientific knowledge involves ‘judgment about things that are universal and necessary,’ as derived from ‘first principles’ (1140b). Practical wisdom, on the other hand, is ‘concerned with things human and things about which it is possible to deliberate’ (1141b). Aristotle is drawing from a tradition in which ethics (in the classical sense) is implemented within the life of the polis and therefore ethics requires politics for its implementation. This ancient link between ethics and politics is less significant in contemporary practical ethics, a theme we will return to in the second part of this paper. Prudent thinkers exhibit different types of excellences because they have different ends in mind. In this respect Aristotle departs from his mentors (Socrates and Plato) when he says that ‘there is no such thing as excellence in practical wisdom’ because ‘excellence’ is a term that involves comparison of particular activities that are variable with respect to time, place, and circumstance, or as the occasion demands (1140b and 1109a).

For Aristotle, ethics and politics belong together because the natural state of mankind is to live in some form of organized society (1169b). A rational agent, for Aristotle, is one who gives serious consideration to a discipline because ‘the man who has been educated in a subject is a good judge of that subject’ (1095a). However, because politics is not an exact science, a rational agent should not expect to gain any more clarity than the subject matter admits (1094b). The central focus of the NE is therefore on types of behavior (virtues) that moral agents ought to exhibit if they intend to live well in a civilized society (polis). By placing the search for the Good within the realm of politics, Aristotle links the flourishing of an individual life to the success of the polis itself. He disputes the separation of ethics from politics, primarily because ethics presupposes a cooperative polis. Politics, in the sense that Aristotle understood it, is the most authoritative science because the purpose of other disciplines is to contribute to the flourishing of the polis itself. Thus, for Aristotle, practical ethics is primarily concerned with social habituation, and therefore the role that reason plays in moral deliberation is not a purpose-independent concept.

Whatever the telos of the whole human life is – and this has proved difficult to define – Aristotle’s cautious approach to unpacking human flourishing (eudaimonia) seems appropriate. For Aristotle, human flourishing begins with a rough sketch of the good life that is filled in with more detail depending on what activity or practice a rational agent engages in (1098a). He

---

An Anthology of Philosophical Studies

seems to have in mind here that prudence or practical wisdom involves skills acquired over time. He illustrates this with the contention that young men can become ‘geometrickians and mathematicians,’ but a ‘young man of practical wisdom cannot be found’ (1142a). The reason for this is that intelligent young men can understand mathematics because mathematics requires knowledge, but not skill. Practical wisdom, on the other hand, requires both knowledge and skill and therefore young men have not had the time to develop the necessary skills associated with a complex social activity. The phrase pros ton kairon, which usually translates as ‘as the occasion demands’ or ‘with an eye to the occasion,’ is used to emphasize the temporal, local, and conditional aspects of practical rationality. A rational decision for Aristotle is one made at the ‘right time,’ with the ‘right motive,’ and in the ‘right way’ (1109a).

Because ethics for Aristotle is a practical rather than a theoretical science he illustrates this in Book III of NE by stating that a rational agent does not spend time deliberating over obvious ends. For instance, a doctor ‘does not deliberate whether he should heal,’ nor an orator ‘whether he shall persuade,’ nor a ‘statesman whether he shall produce law and order’ (1113a). A moral agent ought to be more concerned with practical intelligence or prudence and thus the concentration of thought for doctors, orators, and statesmen is to ‘assume the end’ (health, persuasion, and law and order) and to spend time contemplating ‘how and by what means’ the end ‘is to be attained’ (1113a). A beginner practitioner involved in a long-lived social practice, such as medicine, soon becomes aware that collegial friendships, friendships of utility, are necessary for practical excellence in the discipline concerned. Skill development in medicine initially involves watching others and listening to instruction because the practical skills associated with achieving the end of health takes time to learn. So habituation comes only after one knows what types of actions or behaviors one ought to habituate to achieve the desired end. Similarly, when we do things for fun we soon realize that friends of pleasure enhance the experience of play.

Playing tennis on one’s own might be good for one’s health but playing with a partner is a fun activity as well as being healthy. However, because one can be an excellent doctor or an excellent tennis player, without being a good human being, a morally good human being needs friends of culture or excellence because they supply the psychological and practical honing necessary for shaping a raw human into a person of moral character. Moral character, in the Aristotelian sense, is relationally dependent because no single human being has the capacity or the time to appreciate the internal goods necessary for all activities necessary for a flourishing polis.

The final sentence of the NE implores the reader to make a beginning of the discussion at hand. With this in mind the second part of this paper will argue that the discussion at hand can best be served by acknowledging that Aristotle was right, ethics and politics belong together, but we should not expect our discussion to deliver more precision than it deserves or needs. A type of conversational pragmatism provides the necessary tools of engagement for moral agents who are in genuine disagreement on many issues but who share a
common hope for a flourishing civil society.

Section 2: Pragmatism and the Common Life

For Aristotle, practical ethics is forged in a shared teleological activity of a specific practice (i.e., medicine→health; law→justice; politics→civil society), and the virtues of these practices are shaped by hands-on engagement of the practitioners concerned. Doctors, lawyers, and politicians can and do achieve standards of excellence sufficient to be labeled as experts in their respective fields. This engaged version of practical wisdom stands in stark contrast to the disengaged decision making models currently being advocated in contemporary versions of practical ethics.

Most versions of practical ethics attempt to side-step the conceptual incommensurability of rival moral theories by advocating decision making from the perspective of an ‘impartial spectator’ or ‘ideal observer.’ Peter Singer does this when he follows Sidgwick in claiming that a rational moral agent ought to make decisions from the detached ‘point of view of the universe.’ Singer states that ‘it is a self-evident truth that from “the point of view of the universe,” the good of one individual is of no greater significance than the good of another.’ Thus, as far as Singer is concerned, there is at least one important ethical judgment that is self-evidently true, which is why he does not think of himself a non-cognitivist. In other words, Singer attempts to ground utilitarian impartiality, at least to a limited extend, in a form or moral realism.

Bernard Williams provided an early warning against this detached impartiality when he argued that Sidgwick’s point of view of the universe is nowhere to be found. He suggests that rather than concentrating on a non-existent point of view of the universe, a moral agent should be more concerned with the point of view of here and now and with ‘how a practice hangs together in comparison with other practices.’ This engaged pluralism still seems pragmatically useful because living in a modern polis involves interacting with many people, from a diverse range of social practices, often on a daily basis. For people living in large cities moral agency involves thin agreement over universals and thick agreement over particular practices. The level of human cooperation (engaged pluralism) necessary in a large city is often undervalued. Living in a city with millions of

---

3Singer, Practical Ethics, 334.
other human beings involves a level of cooperation as significant as living in an aboriginal village or a small isolated community of families. The social relationships are generally not as close, but the level of cooperation is as great, or even greater, due to the amount of human interaction necessary when vast numbers of humans live together. In large cities the capacity for human cooperation is often overlooked and is probably masked by media concentration on statistically small levels of violence and other forms of social conflict. Whether or not large human populations are sustainable long term is yet to be determined, but the perception that a larger population leads to more violence and dysfunction is routinely overstated. More significantly, in a large city, a eudemonic telos does not need to be prescriptive because a thin consensus seems all that is required for the polis to function reasonably well.

In a modern pluralist society, a thin level of agreement is evident about many aspects of human action in spite of the lack of agreement over their respective epistemic justifications. Even if this thin level of agreement is restricted to cooperative behavior, and even if this level of agreement was all that could be said about ethics, it remains a significant force for human cooperation. The thin conversational approach to ethics outlined above seems pragmatically more useful, as a minimalist starting point for moral conversation, because it trades on common human teleological imperatives and a common awareness shared by most forms of ethical theory. When thin minimalist claims are combined with thick maximalist practices in a pluralist society, the language of morality provides substantial explanatory power over the moral life because it allows for a social resolution of complex moral issues, even in the midst of serious moral disagreement.

In a pluralist society moral disagreement is precisely what one ought to expect out of moral enquiry because, unlike many other disciplines, moral enquiry has never developed anything like a standard model of enquiry. In science, a standard model is usually an open ended theory or set of laws that provides the epistemological assumptions for further thinking or experimentation. A standard model in science promotes the best explanation hypothesis for a discipline, based on current evidence, but it is also open-ended to incorporate new evidence that can influence the epistemological status of current theories. In a 2002 OECD Global Science Forum the standard model approach was described as ‘a triumph of modern science, with enormous explanatory and predictive power.’\(^1\)

Moral enquiry does not have a standard model primarily because it is not the type of activity that lends itself to this level of precision. Acknowledging that ethics has limits is perhaps the first step in accepting the limits and benefits of moral pluralism. Williams argues for this when he states:

If there is such a thing as the truth about the subject matter of ethics
– the truth, we might say, about the ethical – why is there any

---

expectation that it should be simple? ... Perhaps we need as many concepts to describe it as we find we need, and no fewer.

He goes on to argue that the fact that we appeal to a variety of ethical considerations is precisely what one would expect to find in the complex world we inhabit. Ethical considerations, according to Williams, are ‘genuinely different from one another,’ and this is precisely what moral agents should expect because all of us are ‘heirs to different long and complex ethical traditions, with many different religious and other social strands.’ A useful method for evaluating this diversity is the practice-guided approach advocated by Alasdair MacIntyre. He describes a practice as:

\[
\text{any coherent and complex form of socially established cooperative human activity through which goods internal to that form of activity are realized in the course of trying to achieve those standards of excellence which are appropriate to, and partially definitive of, that form of activity, with the result that human powers to achieve excellence, and human conceptions of the ends and goods involved, are systematically extended.}^3
\]

MacIntyre contends that internal goods of a practice ‘can only be identified and recognized by the experience of participating in the practice in question.’ Without this practical experience one is incompetent to adjudicate which actions or behaviors are internal goods and which are not. The practice-guided approach focuses on a level of agreement that has already been achieved. This is because a practitioner, by virtue of internal standards of excellence within a practice, accepts the authority of those standards in a way that overrides his or her own choices, preferences, or tastes. Practice-guided enquiry does not seek to establish whether actions are right or wrong; rather, it seeks to establish whether actions are consistent with internal goods as defined by the practice. For example, a good medical practitioner is one whose actions and behaviour are consistent with the practice of medicine because a consensus, forged over time, has already been established with respect to actions and behaviours that conform to the telos of medicine (health). The self-authenticating thick standards of a practice focuses on what pragmatists refer to as the cash value or practical consequences for human conduct of decision making because practitioners have access to an established set of internal goods that are historically extended in a manner that is consistent with the teleological imperative of that practice. Conversation within a practice is thick because it involves historically extended argument about the internal goods (virtues) that define it.

---

2Williams, *Ethics and the Limits of Philosophy*, 16.
To enter into a practice is to accept the authority of those standards and the inadequacy of my own performance as judged by them. It is to subject my own attitudes, choices, preferences and tastes to the standards which currently and partially define the practice.\footnote{MacIntyre, After Virtue, 190.}

The self-authenticating aspects of a practice are not absolute because a long-lived practice is continuously refined as better methods of achieving its aims are developed. This alternative approach builds on a thin consensus over practices or activities that promote a flourishing civil society (i.e. health & justice) by recognizing that specific practices (i.e. medicine and law) have their own thick self-authenticating standards of excellence that emerged over time. This approach has similarities to the culture as conversation approach advocated by Richard Rorty.

The notion of culture as conversation, rather than as a structure erected upon foundations, fits well with this hermeneutical notion of knowledge, since getting into a conversation with strangers is, like acquiring a new virtue or skill by imitating models, a matter of σημειωτικής rather than επιστημή.\footnote{Rorty, R. (1979) Philosophy and the Mirror of Nature. Princeton: Princeton University Press, 319.}

For Rorty this type of consensus is not only the best that philosophical objectivity can hope for but also, and more significantly, all that is necessary for moral philosophy to progress. This pragmatic approach overcomes the illusion of foundationalism because it works within established social and political frameworks. These frameworks replace what Rorty calls the ‘notion of knowledge as the assemblage of representations’ with a pragmatist awareness of what people actually do.\footnote{Rorty, Philosophy and the Mirror of Nature, 126.}

Philosophy as conversation can facilitate three different types of conversation that can be usefully employed in a pluralist society. The first type of conversation focuses on thin teleological agreements that people in a pluralist society share – and also that most moral theories endorse – namely, a shared awareness of virtues that sustain a civil community: justice, kindness, peacefulness, civility, beneficence, integrity, respect, etc. This primary conversation unites human beings from different cultures because the discussion is about issues that impact on all people, in every culture. Because the conversation over thin universals transcends the boundaries of culture, it is pragmatically useful because it invites people who are generally different from one another to focus on a requirement they share in common, the flourishing of civil society.

The second type of conversation builds on the first by focusing on those practices within a pluralist society on which there is already substantial agreement. In a stable pluralist society, people disagree over many things:
religion, party politics, sport, and, perhaps most intractably of all, moral issues. In the midst of this disagreement, however, there is also considerable agreement that practices such as medicine, law, education, and politics (in the Aristotelian sense) sustain a civil society. Rorty refers to this type of pragmatism as ‘epistemological behaviorism’ because it invokes knowledge gained from social practices.\(^1\) He further argues that this social justification of belief, following Sellars and Quine, is the normal conversational discourse that happens amongst knowledgeable peers in particular social practices.\(^2\)

The third type of conversation involves a discussion about the internal goods that sustain practices such as medicine, law, education, and politics. Because this type of discussion entails internal goods, it also entails actions, behaviors, rules, protocols, etc., that MacIntyre rightly suggests are ‘historically extended’ and ‘socially embodied’ within specific practices.\(^3\) One result of this commitment to take history and tradition into account, according to MacIntyre, is that ‘human powers to achieve excellence, and human conceptions of the ends and goods involved, are systematically extended.’\(^4\) Conversation such as this is, first and foremost, a conversation among practitioners, but it engages others whenever the internal goods of a practice intersect or clash with those of another.

In a liberal pluralist society, the concept of philosophy as conversation involves dialogue over the things that moral agents hold in common and also dialogue over things that are conceptually incommensurable. Many of the contentious issues in moral debate are incapable of resolution from a moral perspective. A liberal society supports this diversity, but this diversity is not unrestrained because the freedom to be different is not absolute. The rationale for a conversational and practice-based focus for moral enquiry is that the moral space between rival theories of ethics in a pluralist society is not as problematic as is often portrayed. In the same way that cultural difference does not deny common human imperatives, moral difference ought not to deny common moral imperatives. The thin moral universals that human beings share are often lost in the intractability associated with the focus on divisive issues of applied or practical ethics. Acknowledging pluralism in a liberal society involves an acceptance that moral agents can rarely be separated by the legitimacy or otherwise of their respective moral arguments. Serious thinking moral agents hold opposing views on a range of complex moral issues, but they also hold many things in common. Pluralism invokes a conversation about shared common values because cultural and moral difference never completes the separation between one rational agent and another. Pluralism brings with it many benefits and challenges. Learning to live peaceably in the midst of serious disagreement can be a struggle but it is a struggle worth having. Richard Bernstein puts this succinctly when he contends that learning to live

\(^{1}\)Rorty, *Philosophy and the Mirror of Nature*, 176.
\(^{3}\)MacIntyre, *After Virtue*, 222.
with and among people of rival incommensurable traditions is ‘one of the most pressing problems of contemporary life.’

---

CHAPTER ELEVEN

Reconceiving Philosophy as Bodily Discipline: Clues from Master Zhu

John M. Thompson

A textbook several of my colleagues use defines philosophy as “the rational attempt to formulate, understand, and answer fundamental questions.” This definition is pretty good since we generally view philosophy as a mental activity focusing on reasoning (versus physical activities involving lifting, stooping etc.). And certainly philosophy professors engage in the “life of the mind,” a phrase suggesting our professional duties do not include our bodies. Increasingly, however, I find this silence regarding our bodies disquieting, as it seems to assume a mind/body dualism with a corresponding valuing of the former over the latter. I suggest we consider how doing philosophy can include our bodily being. To do so I turn to the writings of the Neo-Confucian thinker Zhu Xi (1130-1200).1 Among his many accomplishments, Zhu devised an education curriculum aimed at transforming a person intellectually, morally and spiritually. What is especially fascinating is how Zhu makes disciplining students’ bodies fundamental to their training. I believe this has intriguing implications for what we actually do as philosophers.

Importance of the Body and its Place in Confucian Society

Following Foucault et al,2 philosophers have recently been recognizing the body as crucial to understanding humanity. Clearly there are various interpretations of the body and we should acknowledge that the body resides in multiple cultural, economic and political dimensions. At the very least, though, I would argue that the body serves as humanity’s material basis, providing our common biological ground as well as the starting point for philosophical reflection (pace Descartes). The body is central to our personal, subjective identity yet exists publically and objectively. Anthropologist Mary Douglas observes that the body furnishes fundamental images for cultural activity, noting, “The body is a complex structure. The functions of its different parts

---

1“Neo-Confucian(ism)” is an awkward Western neologism referring to the revived Confucian tradition that arose during the Song Dynasty (970-1279) in response to the philosophical challenges posed by Daoism and Buddhism.

and their relations afford a source of symbols or other complex structures.”¹ Phenomenologist Maurice Merleau-Ponty describes the body as our “point of view on the world,”² although we should remember that this “window” is part of the world. The body as physical is also finite, having a specific location in space and requiring constant maintenance (feeding, cleaning, etc.). Above all the body is mortal and vulnerable, something we have known seemingly since the advent of our species. Truly, the body compels philosophical attention.

Understanding Zhu Xi’s view of the body requires some background in Chinese and Confucian metaphysics. In general traditional Chinese culture has never posited a mind/body dualism nor an absolute individual/society dichotomy. Rather, the body is enmeshed in a dynamic world wherein all things are comprised of the same “psychophysical stuff,” qi. Humans, while of the same essence as the cosmos, are distinct in being the most sensitive and conscious of creatures because we have the clearest and most finely balanced mixture of qi. The unique status of human beings entails responsibility for helping the cosmos achieve its proper functioning. Moreover, this metaphysical view is informed by socio-political norms promoting familial continuity that manifest in funerary practices, rites of ancestor veneration, and an emphasis on xiao (filial respect). In such a society the body is a precious gift entrusted to us. As the Lushi chunqiu (“The Springs and Autumns of Mister Lu”), an important Chinese text, says, among our “most important tasks [is] to keep intact what Heaven has granted” (1.2).

Confucianism has furthered such views by stressing the ethical dimensions of our bodily being, along with the obligation to develop one’s body through “personal cultivation” (xiu shen). Both Confucius (551-479 BCE) and Mencius (385-312 BCE) emphasize this duty of “self cultivation,” maintaining that it begins with our biological endowment but extends to bettering others. The goal is to use one’s body as a vehicle for achieving universal harmony.³ The Confucian view of the body and its role in developing humanity is summed up in the opening of the Great Learning (Daxue), a central Confucian text: “From the Son of Heaven [the emperor] to ordinary people, all, without exception, should regard cultivating the person (xiu shen) as the root (i.e. their ultimate concern).”⁴

Confucian scholar Tu Weiming argues that xiu shen occurs via socially approved rites. That is, a Confucian education involves a “ritualization of the body” by which individuals learn proper deportment. Such a process has inherently social, public, even cosmic dimensions:

We do not own our bodies; we become our bodies and through that process of becoming we learn to fully realize ourselves as concrete living human beings. Three salient features should be noted here:

1. the body is a vehicle by which we, as Heaven’s co-creators, participate in the great transformation as responsive and responsible agents;
2. the body is an attainment by which we, as beneficiaries of Heaven, Earth, and the myriad things, sustain and enrich nature as filial children and conscientious guardian;
3. the body is a conduit through which we communicate with all modalities of being in order to realize the ultimate meaning of life in ordinary human existence.¹

This standard Neo-Confucian view of the body has dominated East Asian society for centuries. Philosopher Ellen Zhang notes, “Neo-Confucians maintain the endowment of the body as an indispensable qualification for personal and social existence.”² Furthermore, the Neo-Confucian body ultimately includes all things. As Zhangzai (1020-1077), a major influence on Zhu Xi, proclaims in his celebrated Ximing (“Western Inscription”), “Heaven is my father and Earth is my mother, and even such a small creature as I finds an intimate place in their midst. Therefore that which extends throughout the universe I regard as my body and that which directs the universe I consider as my nature.”³

Before looking at Master Zhu’s teachings on the bodily aspects of philosophical education, however, we should note that for Zhu the body was normatively male. Female bodies, by contrast, were the property of the males in their household, who were to protect these valuable resources. Confucian males were to teach and enforce chastity and decorum to keep the female bodies of their clans pure and controlled, thereby promoting familial and social harmony. This marks a major shortcoming in Zhu’s bodily philosophical program.

**Zhu Xi on the Place of the Body in Philosophical Training**

As the “founder” of Neo-Confucianism, Zhu presents Confucian teachings in a metaphysically complex manner, integrating Daoist and Buddhist teachings and practices, and drawing on a wide range of sources. For Zhu, reality is a process whose diverse aspects can be understood in terms of li

---

³De Bary & Bloom, 683.
(“principle,”) and qi, a duality similar to Aristotle’s distinction of form (morphē) and matter (hyle). As with the Aristotelian duality, neither lǐ nor qi exists separately from the other, and it is qi rather than lǐ that provides the dynamic power of the cosmos.

Because of its unstable nature, qi requires discipline and control, much like cultivating a beautiful plant. In practice this means attending to our physical and intellectual dimensions. Zhu shared the earlier Confucian interest in bodily cultivation as a way to promote harmony, looking for guidance from the writings and examples of Cheng Hao (1032-1085) Cheng Yi (1033-1107), two brothers who were preeminent teachers of the Confucian Way. Cheng Hao, the more spiritual of the two, even speaks of “forming one body with all things” (yi wanwu wei yiti) as central to Neo-Confucian training.¹ For the Chengs and Zhu Xi, attention to the body was a matter of course, and they likely drew upon techniques such as we find in contemporary taijiquan (“fist of the Great Ultimate”) and qigong (“working the qi”) aimed at cultivating our qi to its fullest. Zhu addresses this idea explicitly, noting “the objective of learning is to transform this material endowment (qi).”²

As previously indicated, Master Zhu’s understanding of the body and xiù shen are part of a larger project for transforming humanity. To accomplish this goal, Zhu developed an educational curriculum primarily geared towards aristocrats (shi) but open to commoners (min) as well. While Zhu Xi never lays it out systematically in one place we can infer much from his vast corpus of writings. In the following I outline five dimensions of Zhu’s regimen in which the body is central: caring for the body, guidance for proper reading and recitation, methods of meditation, ritual performance, and promotion of virtuous behavior.

Care for the Body

As noted, caring for one’s body is widely viewed as a duty in traditional Chinese society. This entails not only maintaining its health and nourishing it but also mastering it through knowledge, and proper exercise. All of this was important for Zhu Xi. One way to get a sense of how Zhu understands the importance of the body in such education is to see what he says about the “Lesser Learning” (xiao xue) established in antiquity for boys’ initial schooling. For example, Zhu speaks in detail of training in the traditional liú yì (“six arts”) of ritual, music, archery, charioteering, calligraphy and mathematics,³ a list of skills particularly striking for their bodily focus. Elsewhere Zhu says that early learning included instruction in basic physical

chores such as cleaning and sweeping. The “Lesser Learning,” of course, is the basis for engaging in the “Greater Learning” (daxue), the more advanced investigation of reality. Thus for Zhu, having a properly (and consistently) trained body is a prerequisite for philosophical activity.

Moreover, Master Zhu will not allow us to neglect our bodies when we begin the “Greater Learning.” For Zhu, to know something (e.g. a stalk of bamboo) is to know its li. However, fully understanding li entails realizing that principle is not something external but an intimate aspect our own being. Zhu compares such knowing to the alchemical processes of the legendary Daoist masters:

Principle is some separate thing in front of us; rather it’s in our minds. People must discover for themselves that this thing is truly within them, then everything will be okay. Principle can be compared to what the school of inner cultivation calls “lead and mercury, dragon and tiger” – all our present in our own bodies, not external to them.

While it is tempting to focus on Zhu’s statement that principle is “in our minds,” the “mind” of which Zhu speaks (xin, better translated “mind-heart”) does not exist disembodied, just as principle never exists apart from qi. Moreover, Zhu’s imagery is decidedly materialistic, underscoring the physicality of the relationship between principle and our being.

Understandably, Zhu’s training program calls for great effort (a fact that he takes pains to emphasize) and requires a bodily basis. Thus, for example, he chides students for their lack of prior training, saying, “Only if they’re taught to regard inner mental attentiveness as central and to discipline their bodies and minds will they be capable of making the proper effort.” Furthermore, when speaking of the effort such learning entails, Zhu consistently uses analogies of physical labors (poling a boat, climbing a mountain) and direct physical sensation (tasting a piece of fruit) when discussing the knowledge one aims at attaining. Zhu’s analogies are both clear and significant: the “mental” effort of study is best understood as similar to physical exertion. One’s body is the basis for grasping the nature of learning and its actual practice.

Reading and Recitation

Zhu’s learning program is a very literary process that centers on reading and reflecting on texts. Indeed, to this day the Chinese term for learning (xue) has strong textual connotations. The typical term that Zhu uses for the method of study is nian (“reading”). However, nian does not necessarily mean silent reading and often includes reading aloud. It should be obvious even on casual reflection that reading aloud is a very physical activity, requiring a functioning body with various members (eyes, ears, mouth, brain, hands etc.) as well as a proper physical environment (a written text, sufficient light to see, perhaps a desk or table).

That such study encompasses both mind and body is clear throughout Zhu’s

---

1Ibid.
2Gardner, 124.
3Ibid., 93.
admonitions on learning. As he says, “In reading, you want both body and mind to enter into the passage.” Elsewhere Zhu reminds students that the goal of study is not merely intellectual but needs to engage their whole being: “The books you read, the principle you probe, should be embodied in your person.” Again, one’s body is an essential component of the education process, and should concretely express one’s knowledge and insight.

As I note, nian means reading but also reciting; for Zhu reciting texts is necessary to comprehend their message, and Zhu often pairs nian with another verb, du, “to read aloud.” In one of his works, Zhu advises students that reciting sagely words aloud is far more effective than silently pondering them. This has something to do with the effects the spoken sagely words have on one’s qi although Zhu admits to his ignorance of how this works:

The value of a book is in the recitation of it. By reciting it often, we naturally come to understand it. . . There’s value only in recitation, though I don’t know how the mind so naturally becomes harmonious with the psychophysical stuff [qi], feels uplifted and energized, and remembers securely what it reads. Even supposing we were to read through a text thoroughly, pondering over and over in our minds, it wouldn’t be as good as reciting it. If we recite it again and again, in no time the incomprehensible becomes comprehensible and the already comprehensible becomes even more meaningful. . . the value of a text is in the intimate recitation of it. There is no other way.

Clearly, the techniques of reading Zhu espouses are very bodily; for him, effective learning is situated within and through one’s body. Further evidence of Zhu’s view of reading as a bodily practice shows that he understood it could have observable results. In the Jinsilu (“Reflections on Things at Hand”), an anthology of Neo-Confucian texts, Zhu includes the following passage from his personal teacher Cheng Yi:

There are people who have read the Analects without having anything happen to them. There are others who are happy after having understood a sentence or two. There are still others who, having read the book, love it. And there are those who, having read it, “unconsciously dance with their hands and feet.”

At the end of the passage Cheng Yi quotes the preface to the Shijing (“Odes”), perhaps the most prized of Chinese Classics, and Zhu, by including this passage in his anthology, clearly advocates the same idea. For Zhu and his

---

1Gardner, 146.
2Ibid., 143.
3Ibid., 137-139.
Reconceiving Philosophy as Bodily Discipline: Clues from Master Zhu

teacher, reading promotes a joyful unity of mind-heart and body.

In fact, reading and study is so much more than a quiet mental activity for Zhu Xi that he uses physical, even violent imagery when exhorting students in their efforts:

In reading a text, you must be full of vigor. Arouse your spirits, keep your body alert, and don’t let yourself grow weary – as if a sword were at your back. You must pierce through each passage. “Strike the head, the tail responds; strike the tail, the head responds.” . . . Here’s what is necessary: one blow with a club, one scar; one slap on the face, a handful of blood. Your reading of what other people write should be just like this. Don’t be lax!¹

This is truly powerful imagery. It seems clear that Zhu seeks to provoke the sort of all-encompassing response that is essential to true Confucian learning.

Meditation

One particularly interesting aspect of Zhu’s program is his emphasis on jingzuo (“quiet sitting”), a type of seated meditation. This is the most explicitly spiritual aspect of Zhu’s curriculum, as well as the most controversial for its striking Buddhist parallels. Despite such controversy, though, jingzuo forms a vital component of Zhu’s education regimen, so much so that Zhu’s unofficial motto was panzhi jingzuo panzhi dushu (“A half-day of quiet-sitting, a half-day of reciting books”). For Zhu, this practice helps to develop jing (“inner mental attentiveness”) towards all aspects of life, and is a means of balancing tranquility and activity that mirrors the continual back and forth of yinyang, the natural rhythms of the world. Thus “quiet sitting” is not a break from daily life but a valuable method of sagely learning.

Although Zhu rarely specifies actual techniques and postures of “quiet sitting,” he does drop hints every now and then. For Zhu Xi, “quiet-sitting” involves one’s entire being, and is rooted in his understanding of the human body as a delicately balanced coalescence of qi, a living and dynamic mixture of tranquility and activity. He writes, “In the human body there is only a [combination of] activity and tranquility. Tranquility nourishes the root of activity and activity is to put tranquility into action. There is tranquility in activity. For example, when the feelings are aroused and all attain due measure and degree, that is tranquility in activity.”² To the extent that jingzuo (an expression of tranquility if there ever was one) aims at promoting proper activity in the world (i.e. good, refined behavior) it must be integrated into one’s being as a means of guiding one’s interactions with other people and things.

Master Zhu also speaks of jingzuo as a vital complement to a disciplined program of reading and study, specifically as a way of aiding memory and consolidating sagely lessons. In one passage, for instance, Zhu relates the case

¹Gardner, 130.
²Chan, Source Book, 607.
of an earlier thinker who, inspired by reading Mencius, was able to retrieve his “lost mind.” Zhu notes, “Thereupon he shut his door and engaged in quiet-sitting; for more than a hundred days he didn’t look at a book and thereby retrieved his lost mind. He then went to read, and with one glance he absorbed everything.”¹ This passage suggests that jingzuo is a powerful pedagogical technique that can have seemingly miraculous results. Zhu ventures no explanation but most likely it has something to do with the latent abilities released through cultivation of one’s qi. Once more, the bodily aspects of our being are inseparable from our intellectual and spiritual development.

Zhu makes a great show of differentiating “quiet sitting” from the Buddhist practice of “sitting meditation,” suochan (zazen in Japanese). While this reflects his sensitivity to accusations that he and other Neo-Confucians were “closet Buddhists,” Zhu stresses that “quiet sitting’s” primary value is its relation to practical affairs. He says, “In practicing quiet-sitting you mustn’t sit in Ch’an-like contemplation, cutting off all thought. Just gather in the mind; don’t let it run wild with idle thought and it will be at peace and of itself concentrated. Then, when affairs arise, it will respond to them as it should. Once the affairs have passed, it will again be at peace.”² Here Zhu stresses that “quiet sitting” involves great mental discipline, it cannot be understood or practiced in some vague, disembodied state. The point, rather, is to keep one’s mental and physical aspects united in the here and now. Once again, Zhu’s sagely training cannot be divorced from bodily existence.

Ritual Performance

A central theme in Confucianism is the importance of li. Often translated as “rites,” li encompasses more than religious ceremony or court rituals, and includes customary social actions, etiquette and general decorum. Essentially, the li comprise the formal patterned behavior by which one gets along in society. Zhu had a lifelong interest in the study of the li and critically examined them from the time of the antiquity up through his Neo-Confucian predecessors. One of his main concerns was how the ancient ways could be adapted to the present.³ Not coincidentally, many scholars point out that the most distinctive aspect of ritual is its inherently bodily dimension.⁴

Proper performance of the li typically required far more than just superficial politeness. Rather, a cultivated scholar had to master a complex choreography of bodily gestures and expressions along with his intellectual pursuits. Only by learning such things could one have the elegant bearing expected of a shi. This is certainly the case when we see what Master Zhu tells his students. For example, when explaining how to understand and maintain “inner mental attentiveness,” Zhu approvingly quotes Cheng Yi: “Just appreciate fully the flavor of these phrases [of Cheng Yi] – ‘be ordered and

¹Gardner, 145.
²Ibid., 176.
³De Bary and Bloom, 744-745.
solemn,’ ‘be dignified and grave,’ ‘change your countenance,’ ‘set your thoughts in order,’ ‘regulate your dress and dignify your gaze,’ – and make a concrete effort at [doing what they say].”¹ In these aphorisms we see how important specific aspects of behavior and demeanor – facial expression, dress, manner of surveying one’s visual field – are. Clearly, students should strive for a composed and attentive state that is marked by distinctly physical signs.

Most basic training in the li necessarily takes place in the home even before any formal schooling. Zhu understood this fact and took pains to outline proper ritual training for ordinary people in his treatise on family rituals. In the preface to this work he writes:

Ritual “has fundamental elements and elaborations’ [articulated form]. From the perspective of how ritual is carried out in the home, the fundamental elements are to preserve roles and responsibilities and give substantial form to love and respect; the elaborations are the ceremonies and specifications for capping, weddings, funerals, and ancestral rites. The fundamental elements are the daily courtesies of house-holders, the things they must not fail to perform for even a single day. The articulated forms serve further to regulate the beginning and ending of human affairs.²

In this passage Zhu echoes the views of Xunzi (ca. 310-219 BCE), an early Confucian who argues that learning the li is the way to transform humanity from its originally animalistic state. Like Xunzi, Zhu speaks of the li as the fundamental methods for giving order and coherence to human life. And of course, these rites by their nature involve coordinating participants’ bodies.

Master Zhu also provides specific instructions on how to study by cultivating jing, an attitude of careful attention to one’s behavior in the present situation. Much like we see with Zhu’s discussion of “quiet sitting,” so one of the most interesting aspects of such attentiveness is just how bodily it is. Beginning with a quote from the Liji (“Book of Rites”), Zhu writes:

“Sit as though you were impersonating an ancestor, stand as though you were performing a sacrifice.” The head should be upright, the eyes looking straight ahead, the feet steady, the hands respectful, the mouth quiet and composed, the bearing solemn – these are all aspects of inner mental attentiveness.³

This passage prescribes specific, highly ritualized behavior yet these physical actions seem to be inseparable from the mental aspects. For Zhu it seems that proper mental states can only arise from the proper physical postures. Once more, the body is key.

¹Gardner, 171-172.
²De Bary and Bloom, 745.
³Gardner, 172.
Virtuous Behavior

Ethics was a major focus of Confucianism from the beginning. Cultivation of ethical virtue (de) was perhaps the central obligation for Confucians, and the only means for becoming ethical exemplars. Such striving after virtue is, in the Confucian perspective, essential for establishing a society in accord with Dao. In truth, Zhu’s entire project is premised on individual and collective cultivation of virtue. What we should always remember, of course, is that any system of ethics, even when focused on dispositions and attitudes, invariably also focuses on bodies – how they interact with each other and the proper ways they should comport themselves. Once again, Zhu’s views call for careful attention towards and integration of the body in developing and enacting virtuous behavior.

Perhaps the most obvious area wherein we see the body’s centrality to the promotion of virtuous behavior lies in Zhu’s consistent assertion that action is the extension/display of knowledge, a common Neo-Confucian theme. Zhu seems to regard them as mutually dependent upon each other, although he distinguishes them in that while chronologically knowledge comes first, action is more significant. As he says, “When you know something but don’t act on it, your knowledge of it is still superficial.” While Zhu does not equate knowledge and action (as does his later rival Wang Yangming, 1472-1529), he clearly sees knowledge and action as intertwined. Moreover, not physically enacting our knowledge (i.e. not putting into practice the virtues) is “proof” of ignorance; one is, as it were, intellectually and ethically retarded. If we truly have ethical knowledge, we will enact it in daily life.

So important is virtue for Zhu Xi that he lobbied to have moral conduct as a category within the official examination system, the primary means for professional advancement during his time. In “Personal Proposals or Schools and Official Recruitment” Zhu proclaims,

The reason it is necessary to establish moral conduct as an examination category is that virtuous conduct has great importance for humankind. In fact, morality is inherent in human nature and is what is proper to the human Way. It is called virtue [de] because it is inherent [de] in the mind-and-heart. It is called conduct [xing] because it is carried out [xing] by the individual person. [Moral conduct] certainly is not an ornament one puts on in order to please others’ eyes and ears. If scholars truly make an effort in this, they cannot only cultivate their own persons but also extend it to governing others, the state, and all-under-Heaven.

This attests to the interconnection between an individual’s nature, actions, and interaction with others. For Zhu and his students, this is the way to a society exemplifying Dao, a tantalizing prospect whose realization hinges on people embodying ethical principles and behaviors.

1Ibid, 116.
2De Bary & Bloom, 739.
The importance of bodily cultivation for Zhu is especially evident in the articles he drew up for reviving the famous White Deer Grotto Academy. Zhu posted a series of guidelines for academy students as a way of outlining the proper balance between self-cultivation and social relations in ethical and intellectual development.¹ Zhu appended to these articles his personal postscript that makes clear his understanding of what a true Confucian education should be about; “I [Zhu] have observed that the sages and worthies of antiquity taught people to pursue learning with one intention only, which is to make students understand the meaning of moral principle through discussion, so that they can cultivate their own persons and then extend it to others.”² Once more Zhu stresses the connection between our ethical and intellectual aspects, and how our personal development and behavior is bound up with others. For Zhu Xi, training and education shapes one’s person in multiple dimensions of body, speech and mind.

While by no means exhausting Zhu’s work, I have clearly shown the importance of the body in his system. Zhu requires attention to and use of the body as one develops towards becoming a sheng ren. The idea that this process could be a purely “mental” endeavor is unthinkable. Rather, it is a gradual path that proceeds from the outside inward, beginning with our grosser aspects and moving on to finer and ever more subtle levels. The extent to which we regard Zhu’s regimen as philosophical training, we must admit that bodily is integral to it.

**Points for Further Thinking**

Clearly we cannot deny the importance of the body in Master Zhu’s philosophy, and his views on such a central topic warrant our consideration. Zhu’s work calls for us to take bodily being and training seriously, both as a philosophical topic and as an integral aspect of philosophical activity. As Pierre Bourdieu notes in a passage echoing Zhu Xi’s own perspective on the body, “The body believes in what it plays at: it weeps if it mimes grief. It does not represent what it performs, it does not memorize the past, it enacts the past, bringing it back to life. What is ‘learned by the body’ is not something that one has, like knowledge that can be brandished, but something that one is.”³ More recently, theologian Kimerer L. LaMothe has called for greater self-consciousness on the part of scholars of religion on the roles their own bodies play in their work.⁴ At the very least I suggest that the way that Zhu integrates disciplined bodily activity into his methods of learning prompts us to rethink what we are doing as academics.

---

¹Ibid., 742-743.
²Ibid., 744.
Moreover, Zhu’s attention to the body as part of Confucian study, while interesting in and of itself, may promote a more nuanced understanding of Western philosophical history. For example, we can find parallels and points of contact between Zhu Xi’s methods of philosophical training and those espoused in ancient Greece. For instance, in Book III of the Republic, Plato outlines an ideal training for future leaders (derived from the time honored traditions of paideia) that focuses on music, culture and athletics. The body also played a central role in the philosophical activity of medieval thinkers. This is especially the case in monastic traditions where meditative reading (lectio divina) was widely practiced within a community whose life was marked by ritual time and regular observance of individual and communal rites. Perhaps it is really in the Modern era (ironically the era in which conquering the physical world becomes a paramount concern) that Western thinkers overlook the role of the body.

Even more intriguing, Zhu Xi’s views of the body suggest other avenues for cross-cultural philosophical discussion. Could our bodily being point the way to a universal human ethics? Such questions may be unfashionable in some circles but to many of us they have undeniable appeal. Edward Slingerland, for one, concludes his examination of early Chinese philosophies of human action with just such a focus on our common bodily nature, saying, “the body and the physical environment in which it is located . . . shared in all general respects by any member of the species homo sapien, ancient or modern – provides us with a bridge into the experience of ‘the other’.”

Certainly guarding against bodily harm and promoting well-being are prima facie good candidates for universal ethical principles.

Studying the importance of bodily training in Zhu Xi’s work has more immediate relevance as well. Zhu encourages us to take embodiment seriously, stresses our concrete situation in the here and now, and underscores our connection with each other and the cosmos. Such challenges to prevailing views of what it means to be human can have very practical implications. To use just one example, integrating Zhu’s Neo-Confucian views of the body with Western medicine may help develop more successful and humane health care.

Nonetheless, the androcentric views of the body enshrined in Zhu’s philosophical regimen severely restrict the applicability of his pedagogy. Moreover, they sharply contrast with contemporary feminists who seek to put the body back at the center of philosophic discourse. Perhaps, drawing on the age-old image of yinyang, the metaphysical polarities that need each other because of their oppositional natures, our discussion should include some

---


gynocentric analyses of the body to balance our overwhelmingly yang orientation thus far. Regardless, Master Zhu presents a compelling case that a robust philosophical education must include the body as well as the mind.

Bibliography


LaMothe, Kimerer L. “What Bodies Know About Religion and the Study of It,” Journal of the American Academy of Religion 76 (September 2008), 573-601


Modern, 19th and 20th Century
CHAPTER TWELVE

Descartes, Luther, and the Fifth Lateran Council

Aderemi Artis

In the dedicatory epistle to his Meditations, Descartes presents his work as a fulfillment of Apostolici regiminis, the bull produced by the eighth session of the Fifth Lateran Council: ‘the Lateran Council under Leo X, in the eighth session, has expressly commanded Christian philosophers to demonstrate the truth with their utmost strength.’\(^1\) An examination of the text of Apostolici regiminis seems to corroborate Descartes’ claim. We there find the council proclaiming, ‘We command each and every philosopher teaching in universities of general studies or teaching publicly...to use all their efforts to make clear to their students the truth of the Christian religion.’\(^2\) It therefore would have been a bitter irony had Descartes lived to see his Meditations and other works placed on the Index of Prohibited Books in 1663, thirteen years after his death.

The relationship between Descartes’ work and the Christian religion have been the subject of both scholarly and lay attention from the beginning of his philosophical career. The topic of this essay will be confined to the specific issue of the relationship between the Meditations and Apostolici regiminis, although we will have occasion to discuss selected other elements from Descartes’ oeuvre. My primary aim will be to show that when we attend to the specifics of the bull, and to the content and character of Descartes’ claims in the Meditations and elsewhere, we find that Descartes does an exemplary job of carrying out the central doctrinal decree of the bull.

I will begin with an examination of the assumptions made about Apostolici regiminis by much Descartes scholarship, and will argue that the chief doctrinal decree of the bull is to deny the doctrine of double truth and assert there to be a single truth. I then contend that Martin Luther played a significant role in confounding the issue of the immortality of the soul with that of the doctrine of the double truth. Therefore, by the time Descartes began his work, specific doctrinal controversies prompted by the Reformation had sensitized his readers to topics such as the immortality of the soul and transubstantiation, to the

---

\(^1\)‘Concilium Lateranense sub Leone X habitum, sessione VIII, & expresse mandat Christianis Philosophis...veritatem pro viribus probent’ VII, 3. All translations are my own; I have benefited throughout from consultation with various extant English translations, as can be found in the bibliography. All quotations from Descartes are from the Adam and Tannery edition, all quotations from Luther are from the Weimar edition.

\(^2\)‘Insuper omnibus et singulis philosophis in universitatibus studiorum generalium, et alibi publice legentibus...mandamus...eisdem veritatem religionis christianae omni conatu manifestam facere’ 606 in Tanner (1990).
neglect of the overriding theme of the doctrine of the double truth. Finally, I show how, throughout his career, Descartes staunchly held to the notion that there is only one truth, that truth cannot contradict truth, and that philosophy and reason, rightly employed, can never produce claims contrary to the established doctrines of the Holy Catholic Church.

The Significance of Apostolici Regiminis

Descartes’ presentation of his Meditations as a fulfillment of the bull of the Fifth Lateran Council has not gone unnoticed by scholars. Most scholarship that takes up the relationship between the Meditations and Apostolici regiminis proceeds according to two key assumptions: first, that the purpose of Apostolici regiminis is to decree the human soul to be immortal, and, second, that the key relationship between the Meditations and the bull centers on the same topic. In addition to these two key assumptions, some scholars conclude that Descartes’ attempts, while perhaps sincere and valiant, were, in the end, a failure.

An exemplary instance of this approach can be found in Stephen Gaukroger’s Descartes: An Intellectual Biography, where he claims that the purpose of ‘the decree of the Lateran Council of 1513’ was to attack ‘Alexandrian and Averroist heresies,’ and that Descartes in the Meditations hoped to ‘defend the Church orthodoxy on God and the soul.’ Gaukroger gives a succinct summary of the heresies he sees as the target of Apostolici Regiminis: ‘in its Averroistic version...the intellect is in no way personal because mind or soul, lacking any principle of individuation in its own right, cannot be apportioned one to each living human body...in its Alexandrian version...the soul is conceived in purely functional terms; in either case, personal immortality is denied, and its source in both versions is Aristotle himself.’ Finally, with regard to Descartes’ relationship with these heresies, Gaukroger concludes, ‘Alexandrianism and Averroism are the Charybdis and Scylla through which Descartes must steer a passage in setting out his doctrine of the nature of the mind. In this, he faces insuperable difficulties.’

I contend, in contrast to the kind of position taken by Gaukroger and others, that the primary doctrinal goal of Apostolici regiminis is not to decree any specific doctrine regarding the human soul, its immortality, and the metaphysical underpinnings of this immortality, but rather to decree the doctrine of ‘single truth’ against that of ‘double truth.’ The most extreme version of the doctrine of double truth is, essentially, dialetheism – that a

---

1337.
2Ibid.
3Ibid.
4149.
5348. See also C. Fowler, in Descartes on the Human Soul: Philosophy and the Demands of Christian Doctrine, who writes that ‘at the Fifth Lateran Council...[the human soul was] defined as immortal, in opposition to the sixteenth century followers of the Averroist and Alexandrian interpretations of Aristotle’ 312.
6My account of the council’s aims closely follows that of Constant (2002).
Descartes, Luther, and the Fifth Lateran Council

proposition and its negation can both be true. This is the position attacked in the condemnation of 1277, which claims that the doctrine of double truth is that “there is a truth resulting from philosophy, but not catholic faith, as if there were two contrary truths.” However, it is doubtful that any theologians or philosophers targeted by the condemnation actually held such an extreme view as naked dialetheism. A weaker but more plausible version would be that reason can arrive at conclusions in direct opposition to those of faith, which need not imply dialetheism, but is potentially more threatening because it opens up the possibility that faith might have to give way to reason on a given issue. This version will occupy our attention.

The central doctrinal statement of *Apostolici regiminis* is the decree that truth cannot contradict truth, and that the truths and conclusions of reason and philosophy can never come into conflict with the truths of faith. The first and most compelling reason to believe that the doctrinal declaration of the eighth session was directed at the issue of single truth versus double truth, and not at the issue of the immortality of the soul, is a close examination of the text itself. The paradigmatic linguistic expression of formal decree, the verb *decernere* – decree – and the closely related *definere* – define, are found not in the passages that discuss the specific issue of the immortality and nature of the human soul, but in those dedicated to proclaiming the doctrine of single truth against that of double truth. The Council writes that because “truth cannot contradict truth, we define every assertion contrary to the truth illuminated by faith to be completely false...we decree that the detestable and abominable heretics and infidels weakening the Catholic faith will be punished and shunned.” This explains why the disciplinary section of the bull specifies that the seeds of error and heresy are to be found in ‘overly long studies of human philosophy, which God, according to the word of the Apostle, has made foolish and empty, and which, without the spice of divine wisdom and the light of revealed truth, lead into error.” In contrast, when it comes to the passages in *Apostolici regiminis* that take up the issue of the immortality of the soul and its Aristotelian underpinnings, instead of employing the language of decree or definition, the Council rather presents belief in immortality as already established both on grounds of scripture and of a previous conciliar proclamation. The treatment of immortality by the Council, then, is at best merely a restatement of what it took to be existing church doctrine. What the Council seeks to make clear is

---

1“ea esse vera secundum philosophiam, sed non secundum fidem catholicam, quasi sint due contrarie veritates’ 543 in Denifle (1893).
2Pine (1968) argues that Pomponazzi adopts precisely this position.
3“Verum vero minime contradicat, omnem assertionem veritati illuminatae fidei contrariarum, omnino falsam esse definimus...detestabiles et abominabiles haereticos et infideles, catholicam fidem labefactantes, vitandos et puniendos fore decernimus” 605-6 in Tanner (1990).
4“praecipue humanae philosophiae sudia diuturniora, quam Deus, secundum verbum Apostoli, evacuavit et stultam fecit, absque divinae sapientiae condimento, et quae sine revelatae veritatis lumine in errorem...inducunt” 606 in Tanner (1990).
5The biblical passages appealed to are Matthew 10:28 and 25:46, John 12:25, and I Corinthians Corinthians 15:19; the council is that of Vienne.
that the root of doubts on the point of immortality lie in the infestation of ‘human philosophy’ into the field of the divine.

While *Apostolici regiminis* may have begun as an attempt to eliminate the doctrine of double truth in favor of that of single truth, it soon became embroiled in the emerging controversy between Martin Luther and the Roman Church. The issue of immortality rose to the forefront when Luther lampooned the bull as emblematic of the diremption of the church. While Luther at times appears to distinguish between the claims of *Apostolici regiminis* concerning the immortality of the soul and those concerning the doctrine of double truth, he also frequently conflates these two elements of the bull. Thus, while he was in a position to recognize that the core of *Apostolici regiminis* was directed the doctrine of double truth, and that the issues surrounding immortality were seen by the council as growing out of a failure to adhere to the central truth of faith, there are a number of reasons why he nevertheless fails to do so clearly in his published writings. Chief among these are, first, the fact that his remarks on *Apostolici regiminis* are typically launched in a polemical fashion, and, second, that his opinions on the bull were formed simultaneously with worries about how the church had handled, and was handling, the nature of spiritual authority and the proper sources of Christian belief.

**Luther’s Polemics**

*Apostolici regiminis* apparently first came under Luther’s scrutiny in preparation for the Leipzig debate of 1519, and his response to the bull is paradigmatic of the internal dialectic that lasts throughout his career. In the course of his preparations he writes that ‘today everything in the city is utterly wicked and corrupt...what wonder, if under the name of the venerable Roman church so much evil floods into the whole church with full force and vigor, that no one can resist...on account of the study of Roman laws and the neglect of the Gospel, they had to make a statute in the latest Council that the human soul is immortal. What do you think this decree indicates?’ In this instance, Luther connects the material in *Apostolici regiminis* on immortality with the falling away of the church from the revealed biblical word, and, in addition, to the church having become too closely entangled with the secular domain. Thus, on the one hand, he asserts that the Fifth Lateran Council has found it necessary to make a statute on immortality and uses this to ridicule the state of the Roman Church. On the other hand, he tacitly acknowledges that what is really at stake is the broader issue of the primacy of scripture over secular learning. In 1520, not long after Leipzig, he again references the bull, writing that ‘it is even

---

1 Much of this account derives from the work of Headly (1973).
2 ‗Hodie omnia sint in urbe inquinatissima et corruptissima...Quid mirum, si sub venerabili Rhomanae ecclesiae nomine tot mala inudaverint in omnem ecclesiam pleno impetu et praecipiti, cui nemo resister possit...Legum Rhomanarum studium et evangeli neglectus, ut necesse habuerint statuere in concilio novissimo, Animam hominis esse immortalem. Quid, putas, huc decretum indicat?’ II, 226.
agreed by means of Aristotle\textsuperscript{1} and natural reason that the soul is immortal, thus suggesting an appreciation that the relationship between reason and faith is at the heart of Apostolici regiminis. However, in the same year, he writes that it is on account of the ‘worldly splendor’\textsuperscript{2} of the Roman church that it has become necessary to make spiritual immortality an article of faith, intimating that it is simply another symptom of secular corruption that the council had to make pronouncements on anything regarding immortality. Apostolici regiminis returns to Luther’s thoughts periodically throughout his career, and the twin elements of ridicule of the Council’s treatment of immortality based on the perceived worldly corruption of the Church, and recognition that there was a more general issue at stake, vie for primacy in his writing. The fact that the issue of the authority of councils was also addressed by the Fifth Lateran, and that Luther developed charged feelings on this topic in opposition to the council, could only exacerbate his attitude toward its pronouncements considered more broadly.

Luther was not, then, wholly unaware that there were more global issues at stake than immortality for the Fifth Lateran Council when it met in its eighth session. But his outrage and disgust at the Roman church led him to periodically focus on immortality as symptomatic of this corruption. And as Luther’s antipathy to the church of Rome accelerated, it became more and more difficult for him to engage in a measured analysis of conciliar decrees. The emphasis on the making of formal decrees on a host of specific doctrines at the later Council of Trent, along with the fractal processes of doctrinal confessionalization that permeated late sixteenth and early seventeenth Europe, meant that specific topics such as immortality highlighted by Luther became issues upon which lines could be drawn and factions demarcated.

\textbf{Descartes and the Double Truth}

Thus, by the time Descartes began his work, the atmosphere had become charged with a multitude of specific confessional disagreements; nevertheless, there was no point at which Descartes believed the rational enterprise of philosophy he pursued to be incompatible with established and official Catholic doctrine. Indeed, quite the opposite: he was convinced that his own philosophy was superior to that of his competitors on every major front on which his claims intersected with dogmas of Roman Catholicism. He was convinced of this for the precisely same reason that the church was convinced of the ultimate harmony of faith and reason: truth cannot contradict truth. Thus we see Descartes struggling again and again to map out his particular account of a host of specific issues related to church teaching, such as the nature and immortality of the soul, the structure of the solar system, and the Eucharist, and these struggles are present in both his private correspondence and published work spanning many years. In none of these instances does Descartes

\textsuperscript{1}‘Es ist auch beschlossen durch...Aristoteless’ VII, 425.
\textsuperscript{2}‘weltlicher pracht’ VI, 433.
relinquish the overarching principle of single truth and primacy of sound Catholic doctrine. Since the question of the relation between the truth of faith and the truth of reason arises not only in the topic of the immortality of the soul, but also the other particular topics explored by Descartes, it will provide a helpful background to briefly examine his treatment of two of these other topics, the Eucharist and Copernicanism. In addition, we will examine some of the occasions Descartes takes to address the topic of the relationship between faith and reason more generally.

Descartes believed that his philosophy provided a better explanation for the Eucharist than its Aristotelian competition, and tried valiantly to convince interested contemporaries.¹ In a 1638 letter to Vatier he writes, concerning the publication of ‘my physics and metaphysics...fundamentally, I have no fear that anything against the faith will be found there, because, on the contrary, I boast that it has never rested so firmly on human reasons as it does on my principles. More particularly, transubstantiation, which the Calvinists seize on as impossible to explain by the ordinary philosophy, is very easy to explain by mine.’² Similarly, in another letter to Vatier appended to the Latin edition of the Meditations, he gives the Eucharist as an example of how his philosophy surpasses his competition, and in this context writes, ‘since it is never possible for one truth to oppose another, it would be impious to fear that truths discovered in philosophy could oppose those of faith.’³ The parallel to the bull Apostolici regiminis is unmistakable. Copernicanism, like the Eucharist, was an issue that vexed Descartes for many years, and although his attitude toward it was somewhat different from that of the Eucharist, his conviction of the compatibility of faith and reason never wavered. Much of Descartes’ thoughts on the issue can be surveyed in a series of letters, most of which were written to Mersenne in a period spanning over a decade from 1633 to 1644. In an early instance revealed in a 1633 letter, Descartes discusses the status of the Copernican system, and the relationship of his Le Monde to this system. It is clear that Descartes had been, during the composition of Le Monde, convinced of Copernicanism, and had believed it to be compatible with Catholic doctrine. However, in his letter he says that he has since become aware that important authorities in the Church have come down against Copernicanism, and writes, ‘I admit that if it is false, all the foundations of my philosophy are also false, because it can obviously be demonstrated from them...but as I would not, for anything in the world, want something to come from my one discourse in which the smallest word could be found that had been disapproved of by the

---

¹For in-depth discussions see chapter seven in Ariew (1999), Armogathe (1977), and Gouhier (1972).
²‘Ma Physique & Metaphysique...Je crains nullement au fons qu’il s’y trouve rien contre la foi; car au contraire j’ose me vanter que jamais elle n’a estre si forst appuyee par les raisons humaines, qu’elle peut estre si l’on suit mes principes; & particulierement la Transubstantiation, que les Calvinistes reprennent comme impossible a expliquer par la Philosophie ordinaire, est tres-facile par la mienne’ I, 564-5.
³‘Cum una veritas alteri adversari nunquam possit, esset impietas timere, ne veritates in Philosophia inventae iis quae sunt de fide adversentur’ VII, 581.
Church, I also strongly want to suppress it.¹¹ In this case, Descartes’ desire to publicly pursue a particular doctrine concerning the solar system is overridden by the fact that, at that moment, the Church appeared to be leaning against Copernicanism, even if only well below the conciliar and papal levels.² And yet all this is not to say that Descartes has no confidence in reason, because when he is convinced that nothing in his work conflicts with Catholic teaching, as he is when he writes to Mersenne concerning the movement of the earth in 1640, he again closely and strikingly echoes the position of the Fifth Lateran: ‘believing very strongly in the infallibility of the Church, and also not doubting my reasons, I cannot fear that one truth is contrary to the other.’³

We find a similar situation in his general statements about reason and faith. The Principles of Philosophy (1644), Descartes’ last and in many ways most comprehensive major treatise, presents a few brief but paradigmatic examples of his conviction regarding the single truth and the position of supremacy of the universal church. Here he writes, ‘the highest rule to be set in our memory should be that those things revealed to us by God are to be believed with the greatest certainty of all.’⁴ Later in the work, after delivering to the reader a comprehensive compendium of his philosophy, he declares that ‘remembering my frailty, I affirm nothing, but submit everything here...to the authority of the Catholic Church.’⁵ In a polemic written near the end of his life, Descartes once again addresses the question of double or single truth directly. He writes that, while there are some propositions which ‘are believed by faith alone, such as those concerning the mystery of the Incarnation, the Trinity, and such like,’⁶ there are others which, ‘while they pertain to faith, can also be inquired into by natural reason...[such as] the existence of God and the distinction between the human soul and body.’⁷ Here Descartes explicitly offers the existence of God and the distinction between the soul and the body as instances of the more general thesis of the unity of truth, be it arrived at by faith or by natural reason.

It is against the background of Descartes’ conviction of the compatibility of all truth, and of the superior harmony of his own philosophy with Catholic

¹¹‘Je confesse que s’il est faux, tous les fondemens de ma Philosophie le sont aussi, car il se demonstre par eux evidemment...Mais comme ie ne voudrois pour rien du monde qu’il sortit de moy vn discours, où il se trouvast le moindre mot qui fust desaprouvé de l’Eglise, aussi aymé-je mieux le supprimer’ I, 271. It is not without irony that Copernicus himself, in the preface to his De revolutionibus orbium coelestium, positioned his efforts as carrying out the edicts of the Fifth Lateran to amend the ecclesiastical calendar.
²‘Whether or not Descartes ever ceased to subscribe personally to the movement of the earth is beyond the scope of this essay.’
³‘croyant tres-fermement l’infaillibilité de l’Eglise, & ne doutant point aussi de mes raisons, je ne puis craindre qu’une verité soit contraire à l’autre’ III, 259.
⁴‘memoriae nostræ pro summâ regulâ est infigendum, ea quæ nobis à Deo revelata sunt, ut omnium certissima esse credenda’ VIIIA, 39.
⁵‘Memor meae tenuitatis, nihil affirmo, sed haec omnia, tum Ecclesiae Catholicae auctoritati...submitto’ ibid., 329.
⁶‘solâ fide creduntur, quales sunt de mysterio Incarnationis, de Trinitate, & similibus’ VIIIIB 353.
⁷‘quamvis ad fidem pertineant, ratione tamen naturali quæri etiam possunt, inter quas Dei existentia & humanæ animæ à corpore distinctio’ ibid.
doctrinal over competitors, that his dedicatory epistle in the *Meditations* must be read. Moreover, correctly parsing his epistle is not always a simple matter, in part because he often addresses numerous issues in a single thought. Such is the case when he brings up the Fifth Lateran, as is evident when we return to the passage mentioned at the very beginning of this essay. Here Descartes mentions and distinguishes between the specific issue of immortality and the general issue of double truth, and yet fails to keep the two issues sufficiently demarcated. This becomes evident when we look at the statement in fuller context: ‘With regard to the soul, many judge its nature to be far from easy to investigate, and some have even dared to say that human reasons lead to the conclusion that the soul perishes with the death of the body, and that the contrary is grasped by faith alone. However, since the Lateran Council under Leo X, in the eighth session, has condemned them, and expressly commanded Christian Philosophers to refute their arguments and demonstrate the truth with their utmost strength, I have accordingly not doubted to undertake this.’ Here Descartes claims that, while some have held that reason can arrive at conclusions contrary to those of faith, he opposes this in the spirit of *Apostolici regiminis*. However, he also merges this claim with his opposition to a specific position in the debate about the possibility of a philosophical demonstration of immortality, thus making it difficult to recognize that there are indeed two distinct issues at stake. A somewhat similar challenge is presented in another passage in the dedicatory epistle where he writes, ‘although it is enough for we faithful to believe that God exists and that the human soul is not the same as the body, it is obvious that the infidels can be persuaded of no religion or any moral virtue unless they have already accepted these two claims through natural reason.’ Here Descartes indicates that the single truth that God exists, or that the soul is distinct from the body, can be arrived at through faith or through reason, and that while Christians like himself can rely on faith, it would not be possible to convince a non-Christian unbeliever through faith, and thus reason must be employed. However, reason must arrive at the same truth as that arrived at by Christians through faith. Once again, he distinguishes between the general notion that faith and reason must arrive at identical truth, and the specific issues of immortality and God’s existence, and yet again joins their discussion in a way that makes it difficult to clearly and distinctly recognize the difference. Notwithstanding this difficulty, the specific issues treated in the *Meditations* are offered as key instances of the more general principle: this is precisely the position of *Apostolici regiminis*.

1‘quantum ad animum, etsi multi ejus naturam non facile investigari posse judicarint, & nonnulli etiam dicere ausi sint rationes humanas persuadere illam simul cum corpore interire, solaque fide contrarium tenei, quia tamen hos condemnat Concilium Lateranense sub Leone 10 habitum, sessione 8, & expresse mandat Christianis Philosophis ut eorum argumenta dissolvant, & veritatem pro viribus probent, hoc etiam agredii non dubitavi’ VII, 2-3.

2‘Quamvis nobis fidelibus animam humanam cum corpore non iterire, Deumque existere, fide credere sufficiat, certe infidelibus nulla religio, nec fere etiam ulla moralis virtus, videtur posse persuaderi, nisi prius illis ista duo ratione naturali probentur’ ibid., 1-2.
Conclusion

Throughout much of his career, Descartes attempted to show how a host of particular theological truths could be arrived at philosophically. He was guided by the conviction that reason and faith must be compatible because there is only a single truth. In the Meditations, Descartes presents his ideas as the fulfillment of Apostolici regiminis, and insofar as the chief doctrinal aim of Apostolici regiminis is to assert the univocity of truth in opposition to any doctrine of double truth, Descartes is completely in line with the church. However, between the conclusion of the Fifth Lateran Council in 1514 and the publication of the Meditations in 1641, Martin Luther had begun a rebellion against the church, specifically targeting Apostolici regiminis for his vitriolic fire. The Council of Trent was also convened between the Fifth Lateran and the Meditations, leading to a focus on doctrinal specifics and away from the general issue of the compatibility of reason and faith. These events help account for why the efforts of the Fifth Lateran on the single versus double truth have been obscured from Descartes’ time to our own. When we return to the original statement of the council, we can see the doctrine of double truth as its real target. At least with respect to Apostolici regiminis and the spirit it embodied, the Roman Catholic Church erred in its prohibition of Descartes’ works.

Bibliography


This paper looks at Schopenhauer’s critique of Kant in the broad context of their respective philosophical projects. This broad context is needed to do justice to Schopenhauer’s critique, because his attacks on specific elements of Kant’s philosophy often seem, when taken in isolation, to simply reject out of hand various commitments fundamental to Kantian transcendental idealism. As such, Schopenhauer’s critique seems to lack any internal traction in the Kantian system. Particularly striking is Schopenhauer’s refusal, throughout his critique of Kant’s theoretical philosophy, to so much as acknowledge the distinction, vital to Kant, between the transcendental and empirical, or causal, orders. This might well leave the reader suspecting that Schopenhauer fundamentally misunderstands the original and defining elements of Kant’s transcendental idealism.

While this suspicion is not entirely without justification, I will argue that much of Schopenhauer’s broader critique rests on an important, though largely implicit, critique of the conception of the inclinations Kant develops in his account of motivation. In the Critique of Practical Reason, Kant uses the inclinations to coordinate two conceptions of freedom: first, freedom from determination by phenomenal causality (a conception secured ‘negatively’ in the first Critique), and second, freedom from determination by sensible conditions, most notably the inclinations. In brief, Kant effects their coordination by claiming that if the inclinations were our only source of motivation, human agency would be reduced to mere mechanism.\(^1\) This raises a number of difficulties for Kant that I will discuss briefly in the third section below. When we bring together the theoretical and practical aspects of Schopenhauer’s critique in light of these difficulties, we see that he in fact highlights serious internal questions for Kant’s philosophy.

Overview of Kant’s Critical Project and the Place of the Inclinations

Here I can give only a very brief overview of those elements of Kant’s critical project relevant to the Schopenhauerean critique examined below. These elements centre around two key features of Kant’s critical philosophy: 1.

\(^1\)See, e.g., KpV 5:47-49, 5:39, 5:97. Translations given in this paper of KpV are based on Gregor (1996); translations are mine for all other works.
the separation of rational and empirical-causal orders, and 2. the coordination between practical and theoretical philosophy. I address the first of these now, and the second after looking at Schopenhauer’s critique of Kant’s theoretical philosophy.

In his theoretical philosophy, Kant looks to reconcile the competing claims of rationalism and empiricism by carefully distinguishing the spontaneity of the understanding from the receptivity of intuition, and giving to each an essential role in our cognition of the world. According to Kant, the understanding is spontaneous and lawgiving, but only for those objects that can be given in experience, i.e., through intuition, which is to say only for objects taken as appearances, or phenomena. Spontaneity here means, in part, that the activity of the understanding does not itself take place within the phenomenal realm, since all phenomena are, as such, thoroughly conditioned by preceding phenomenal causes and therefore not spontaneous. The lawgiving of the understanding, in the acts of synthesis through which it structures the manifold given in intuition, is grounded ultimately in what Kant calls the ‘transcendental unity of apperception,’1 or the ‘I think’ that ‘must be able to accompany all my representations.’2 Kant is careful to keep this transcendental I distinct from the empirical ego and from anything else that might be given in intuition and thus located in the phenomenal realm.

Similarly, on the practical side, Kant contrasts the spontaneity of reason with the receptivity of feeling. Each of us is, as a rational being (or ‘noumenal self’), both the source of, and subject to, the moral law. This lawgiving of practical reason, unlike that of the understanding, does not depend on intuition to supply the ‘material’ it governs, since practical reason is lawgiving, Kant argues, for things-in-themselves, namely for the actions of rational beings as they are in-themselves, not as they appear in the phenomenal realm.3 However, the actions commanded by the moral law are of course meant to have effect in the phenomenal, empirical world of experience. It is precisely on account of our empirical nature that the commands of the moral law are duties for us, as opposed to rules that infallibly govern our actions. For our empirical nature involves, through the receptivity to feelings of pleasure and displeasure, inclinations which are often in tension with the moral law and which can incite us to adopt maxims of action that conflict with the moral law. When incited in this way, it is our duty as rational beings to restrain the inclinations by determining the will through practical reason alone.

Now Kant clearly does not intend the transcendental I and the noumenal self as entities distinct from the empirical self. They are simply distinct capacities (Vermögen) that belong to each of us as finite rational beings. The point is that the exercise of these capacities – the spontaneous activity of practical reason and of the understanding – is not conditioned in the way phenomena are, i.e., causally. Their exercise does not take place, strictly speaking, within the phenomenal realm, and in particular does not consist of

---

1KrV A108. Often the ‘synthetic unity of apperception’; e.g., KrV B131-135.
2KrV B131-132.
3KU Preface, A/B xviii.
empirical-causal events. Kant insists, in this way, on a distinction between rational activity and empirical nature: on the theoretical side, between, roughly speaking, the transcendental and the empirical-causal orders, and on the practical side, between the rational (or ‘noumenal’) and the empirical aspects of the individual.

I want to take a look at Schopenhauer's critique of Kant’s theoretical philosophy, before addressing the question that naturally arises here, as to how the lawgivings of theoretical and practical reason, respectively, are related.

**Schopenhauer’s Critique of Kant’s Theoretical Philosophy**

Schopenhauer directs much of his criticism of Kant at the thing-in-itself. He does not, however, want to rid philosophy of the thing-in-itself; indeed the ‘great advance’ beyond Kant on which Schopenhauer prides himself is precisely the *identification* of the thing-in-itself with Will. Schopenhauer objects rather to the way in which Kant introduces the thing-in-itself into his system:

Kant bases the assumption of the thing in itself, though hidden beneath a variety of expressions, on a conclusion in accordance with the law of causality, namely that empirical intuition, more precisely the sensation in our sense organs from which it arises, must have an external cause. And yet, according to his own and correct discovery, the law of causality is known to us *a priori*, is therefore a function of our intellect and thus of subjective origin; moreover sensation itself, to which we are here applying the law of causality, is undeniably subjective; and finally even space, in which we locate, through this application [of the law of causality] the cause of sensation as object, is a form of our intellect given a priori, and therefore subjective.¹

Thus the ‘entire empirical intuition’ remains ‘on subjective ground and territory, as merely a process in us, and nothing entirely distinct from it, independent from it, can be dragged in as thing in itself or justified as a necessary presupposition.’²

Striking to a Kantian reader is the causal sense that Schopenhauer gives to intuition. Indeed, Schopenhauer explicitly identifies the intellectual operations of cognition as a whole with causal psychological or physiological processes, interpreting Kantian transcendental acts of synthesis as ‘brain functions’ (‘Gehirnfunktionen’).³ In other words, from a Kantian point of view,

---

¹WWV, p.516.
²Ibid.
³E.g., WWV, p.495: ‘Thus Locke had withdrawn from appearance the part played by the sense organs; Kant now withdrew the part played by the brain functions (although not under this name) […]’ The parenthesis is certainly worth noting. See also p.499, pp.535-6.
Schopenhauer seems to confuse the transcendental with the empirical, that is, he fails to distinguish between the rational and empirical-causal orders.

From Schopenhauer’s perspective, however, Kant is equivocating on the sense in which objects are ‘given’ to us. Schopenhauer points out that if any objects are given to us, these must be the empirical objects we actually cognize, and these affect our sense organs causally, in the unmysterious sense of empirical, i.e., phenomenal, causality. For Schopenhauer, it is precisely through these empirical objects and this causal affection that we have perceptions. Yet Kant insists we need to think a supersensible substrate underlying the ‘giving’ of objects of experience. Again, this is because the manifold of intuition given to the understanding cannot, for Kant, be given in a merely empirical-causal sense. Granted, Kant cannot explain this non-causal sense of ‘giving,’ since, after all, we cannot know how the supersensible underlies the objects of experience. But since the distinction between the transcendental and the empirical is so fundamental to Kant’s transcendental idealism, it seems rather unenlightening for Schopenhauer to simply dismiss it, or fail to see it altogether.

Moreover, we might press Schopenhauer on his own ‘discovery’ of the thing-in-itself as Will. Since Schopenhauer himself posits an underlying reality, of which the empirical world is merely the representation, why is he so quick to insist that Kant’s ‘supersensible substrate’ is meant as underlying cause, in a way that so obviously violates Kant’s own idealist standpoint? While Schopenhauer might not always seem entirely clear on this himself, the answer is that his issue is not fundamentally with the Kantian object, but rather with the Kantian subject. Schopenhauer thinks that the approach Kant takes in the Transcendental Analytic is fundamentally mistaken in isolating the understanding from the object of cognition, in order to derive categories for thinking ‘objects in general,’ i.e., independently of our particular spatio-temporal forms of sensible intuition. This approach leaves Kant putting forward, somewhat matter-of-factly, a table of categories never seriously derived, as to their form and number, from their supposed ground in the transcendental unity of apperception. Furthermore, Kant then needs a second faculty, the transcendental imagination, to produce what he calls spatio-temporal ‘schemata’ of the categories, in order to explain our actual cognition of objects. Schopenhauer argues that whatever plausibility the categories have comes precisely from their role in the actual cognition of spatio-temporal objects, and that the categories could therefore never be anything other than abstractions from actual experience. In other words, the Kantian hierarchy is, according to Schopenhauer, ‘diametrically opposed to our presentation, according to which the concept obtains its value and truth from intuition alone.

Schopenhauer is right that this need not undermine the a priori nature of

---

1 See, e.g., KpV 5:6, KU Preface, A/B xviii-xix.
2 See, e.g., WWV, pp.523-526. I'll leave unexplored here the difficulties inherent in Schopenhauer’s ‘abstractionist’ account of concept-formation.
3 WWV, p.524.
the categories’ application to objects of experience. Schopenhauer is not arguing that the law of causality is established empirically, from experience. The understanding’s application of the law of causality to perception is what first makes experience possible, as a ‘real thing’ in contrast to mere ‘figment of fantasy.’ Similarly, for Kant, the schemata first make experience possible as acts of synthesis in accordance with necessary and universal rules (e.g., every cause must precede its effect in time). In contrast to the categories of the understanding, however, the schemata are specific to our spatio-temporal forms of intuition. Schopenhauer is simply arguing that there is no reason to imagine that the necessary and universal rules embodied in the schemata, for the cognition of spatio-temporal objects, are merely specific instances of rules for the presentation of ‘objects in general.’

Indeed, Schopenhauer rejects the complicated relation of categories and schemata altogether, reducing these to the single law of causality, which does not belong so much to the subject, or to a transcendental understanding, as to the act of perception in which the subject and object are mutually conditioning. The motto of idealism, as Schopenhauer understands it, is: No object without a subject, and no subject without an object. It is really in holding to the second half of this motto that Schopenhauer most clearly opposes himself to Kant, yet he often focuses on the first half, making his arguments somewhat oblique. Of course, it is crucial to Kant’s philosophy as a whole not to take as starting point the mutual conditioning of the subject and object. Particularly for his practical philosophy, Kant requires that the spontaneity of our higher cognitive faculties not be necessarily restricted to their relation to objects of experience.

In a nutshell, this is the core of Schopenhauer’s critique of Kant’s theoretical philosophy: Kant’s initial isolation of the subject from the object, in the deduction of the categories, makes it impossible to maintain a properly idealist standpoint in subsequently accounting for the ‘givenness’ of objects to this subject. For Kantian cognition then requires confronting this independent subject with an equally independent object, i.e., with the thing-in-itself. The strength of Schopenhauer’s critique lies in calling attention to the somewhat mysterious appearance of Kant’s twelve categories, and to the correspondingly mysterious theoretical role of the Kantian thing-in-itself, the mysteriousness of which only deepens once Kant finds a very different practical role for the thing-in-itself.

Nonetheless, Schopenhauer’s critique as considered thus far may seem to offer at best a competing conception of idealism – one not based on a distinction between rational and empirical orders – rather than a critique that genuinely engages the Kantian position. In order to see how Schopenhauer finds an internal toehold for his critique, we need to look at the relation Kant describes between theoretical and practical reason, and then at the connection between Schopenhauer’s critique of Kant’s practical philosophy and his critique of the theoretical just considered.

---

1WWV, p.528.
Kantian Theoretical and Practical Reason: 
Non-infringement and the Role of the Inclinations

How, then, does Kant relate the theoretical and practical lawgivings? He argues, first, that they are established on one and the same ‘territory’ (Boden, territorium), consisting of the ‘totality of objects of all possible experience.’¹ Now, Kant is committed both to the universal determinism of phenomenal causality in the realm (Gebiet) of the understanding’s lawgiving, and to the independence from determination by phenomenal causality of practical reason’s lawgiving.² Kant therefore adopts a doctrine of mutual non-infringement, in which the understanding is lawgiving for objects taken as appearances, while practical reason is lawgiving for things-in-themselves, that is, for the actions, as they are ‘in-themselves,’ of rational beings. The division is summed up in Kant’s statement that ‘understanding and reason thus have two distinct lawgivings on one and the same territory of experience, such that neither may infringe upon the other.’³

In this way, Kant paints a tidy, if abstract, picture of the relationship between the lawgiving of practical reason and that of the understanding. Moreover, the ‘space left vacant’ in the theoretical philosophy – namely the role of the thing-in-itself, necessarily placed by theoretical reason at the ‘basis’ of appearances,⁴ though theoretically unknowable – is filled, in the practical philosophy by the noumenal causality of practical reason’s lawgiving.

I want to raise two points here. First, the very tidiness of mutual non-infringement threatens to reorder the lawgiving of practical reason superfluous, as Schopenhauer is more than happy to charge, labeling the categorical imperative a ‘sceptre of wooden iron.’⁵ Moreover, the tidiness itself rests on Kant’s claim that the thing-in-itself has a purely negative role in theoretical cognition; yet its role as a ‘supersensible basis’ of appearances does not seem so purely negative, nor does it square cleanly with the positive role that the practical philosophy gives to the thing-in-itself, i.e., as rational self-determination. Second, Kant of course needs to supplement this tidy but abstract picture with a more concrete account of human agency and motivation. The account he gives in the second Critique creates serious tension with the doctrine of mutual non-infringement.

The difficulties revolve around Kant’s conception of the inclinations in the second Critique, where they play a key coordinating role between Kant’s account of motivation and his claims of non-infringement between noumenal and phenomenal causality. For Kant, the inclinations belong to our sensible, empirical nature and thus do not reflect our moral characters; they merely reflect our respective tendencies to be affected by certain activities and objects

¹KU Preface, A/B xvii.
²E.g., KpV 5:43; see also KpV 5:94-100.
³KU Preface, A/B xviii.
⁵WWV, p.620.
in pleasurable or unpleasurable ways. They supply the material ends of all our actions, and together constitute our ‘system of happiness.’ Kant claims that if the inclinations were our only source of motivation, then human agency would be reduced to ‘mere mechanism,’ completely determined by natural necessity, i.e., by phenomenal causality. Now, mapping, in this way, motivation through the inclinations to phenomenal causality, and will-determination by pure practical reason to noumenal causality, suggests, in light of the above doctrine of mutual non-infringement, that these two aspects of our agency do not interfere with one another.

But does non-infringement really make sense when translated into Kant’s account of motivation? Let’s consider the relation Kant describes between the inclinations and practical reason. The moral law commands us, generally speaking, to restrain our pursuit of happiness so that it is conformable with the same pursuit by all other rational beings. Crucially, this means that in cases of conflict between the material ends of the inclinations and the commands of the moral law, recognition of the latter ‘infringes upon all the inclinations in self-love’ and ‘humiliates’ our self-conceit. While this humiliation is painful, it has also a positive side as the feeling of respect for the moral law, through which we rise above mere self-love. This feeling of respect is the empirical manifestation, or effect, of pure practical reason’s determination of the will independently of all inclination.

Kant famously illustrates this account of moral motivation with the ‘gallows example’ in which a prince orders the hypothetical agent, on pain of death, to give false testimony against someone whom the prince wants to execute on a suitable pretext. Kant argues that the agent will surely be aware of his own capacity to act in accordance with what duty here commands as his moral obligation. That is, he feels respect for the moral law, as the law of his own rational being, whether or not he ultimately summons the fortitude to act from duty against ‘the love of life,’ which circumscribes all inclinations. The feeling of respect is itself only explicable on the assumption of a source of motivation distinct from all inclination.

Kant is not arguing that to act morally is necessarily to act against one’s inclinations; he does insist, however, that this must at least sometimes be the case. At the end of the Dialectic of Practical Reason, Kant argues that it is fortunate we cannot have theoretical certainty of God’s existence and of the soul’s immortality, for otherwise fear of God and hope of eternal reward would preclude any conflict between our inclinations and the demands of the moral law (thus reducing our actions to ‘mere mechanism’). But it is only through ‘the conflict that the moral disposition now has to carry on with the inclinations’ that ‘moral strength of the soul is gradually to be acquired.’ In the absence of such conflict, ‘reason would have no need to work itself up so as to

---

1 See, e.g., *KpV*, 5:74.
4 *KpV*, 5:74.
gather strength to resist the inclinations,’ in which case ‘the moral worth of actions, on which alone in the eyes of supreme wisdom the worth of the person and even that of the world depends, would not exist at all.’

This necessary conflict between the ‘moral disposition’ and the inclinations sits at the very least uneasily with Kant’s claim of mutual non-infringement between noumenal and phenomenal causality and his ‘placing,’ as it were, of the inclinations within the realm of phenomenal causality. Indeed, the separation between rational and causal orders, so carefully maintained in the theoretical philosophy, is not so easily drawn in the practical. Schopenhauer’s account of motivation holds less ambiguously, as I’ll now explain, to a deterministic empirical-causal view of agency, dismissing the notion of noumenal causality altogether. Schopenhauer thus remains arguably more consistent than Kant across theoretical and practical aspects of philosophy, as far as the supposed distinction between rational and causal orders is concerned.

**The Causal Efficacy of Pure Reason:**
**Infringement on the Phenomenal Realm?**

In this final section, I relate the fundamental differences in the Kantian and Schopenhauerean views of human agency to Schopenhauer’s theoretical critique considered earlier.

At the heart of practical philosophy, Kant puts the question whether pure reason can of itself be practical, that is, whether reason can determine the will independently of the inclinations, or, what is supposed to amount to the same thing, independently of phenomenal causality. Kant’s affirmative answer makes essential use of the categories of the understanding as categories for thinking ‘objects in general’; in particular, that there is at least no contradiction in thinking a supersensible use of these categories, opens the way for practical reason to secure a ‘supersensible object of the category of causality,’ namely freedom – though only for ‘practical use.’

Once we see that Schopenhauer’s critique of Kant’s theoretical philosophy is primarily directed at the Kantian subject, it is a short step to central issues of practical philosophy. In brief, from Schopenhauer’s perspective, Kant first wrongly places the activity of the understanding in a transcendental realm isolated from actual objects of cognition, subsequently uses the categories derived thereby to illegitimately think supersensible objects, and finally has practical reason fill in this ‘noumenal realm’ with a fictitious supersensible kingdom of free rational agents.

As noted earlier, for Schopenhauer, concepts and ‘categories’ can only arise by abstraction from concrete perception, and they therefore have validity only in reference to such perception. Reason can therefore do nothing more than mirror reality: philosophy, for Schopenhauer, ‘is to be a thoroughgoing

---

1*KpV, 5:147.
2See, e.g., *KpV, 5:6.*
Our being as agents, in particular, cannot be constructed out of such abstractions. For Schopenhauer, at least part of what constitutes us as agents must be of a completely different nature from the perception and use of concepts that characterize us as knowers.

Schopenhauer thus argues that our nature as individuals is composed of two heterogeneous elements: that of a knowing subject, and that of a willing subject. It is feeling that characterizes the latter, and feeling is an immediate modification of the will – immediate, in that, unlike modifications of the knowing subject, feeling is not mediated by any relationship to an object. Now, underlying all modifications of will, i.e., all feelings, is one and the same motivational force: the insatiable striving that characterizes the will-in-itself. Whatever the explicit motive, every desire is ultimately an expression of this striving: ‘The motive stands everywhere before the will as a multiform Proteus: it always promises complete satisfaction, quenching of the will's thirst.’

This insatiable striving for satisfaction is, however, manifested differently in different characters. Each person has a unique character that determines which motives move her to which actions; this character is a causal law, binding motives to actions, in exactly the same sense as, e.g., the law of gravity binds gravitational causes to their effects. Schopenhauer, in fact, explains motivation as ‘causality seen from within,’ so that the feeling of desire, or striving, that binds motives to actions is simply the ‘inner’ perspective on causal necessity. Schopenhauer thus attributes will, at least by analogy, to all phenomena, conceiving the will’s striving as the very glue of the causal necessity that reigns throughout the empirical world. For animals – including humans – perception is the medium through which motives bring about action, though, for humans, reason expands this medium through the calculation of consequences not given in immediate perception. In this way, the subjection of the understanding and reason to the will’s incessant striving constitutes us as agents.

Now, Schopenhauer, like Kant, holds to a universal causal determinism throughout the empirical world. Schopenhauer’s conception of agency fits smoothly with this determinism, since he takes the character governing any individual’s actions on exactly the same model as other laws of nature. Agency, being motivated to particular actions, is the inner perspective on causal determination, in a sense simply ‘what it feels like’ to be causally determined.

Schopenhauer’s relating of rational and causal orders is here thoroughly consistent. The rational is always assimilated, or subordinated, to the causal, whether as mirror of the world in cognition, or as medium through which causality acts in motivation. In both cases, Schopenhauer interprets the activity

\[\text{recapitulation, a reflection, so to speak, of the world in abstract concepts.'}^{1}\]

\[\text{Our being as agents, in particular, cannot be constructed out of such abstractions. For Schopenhauer, at least part of what constitutes us as agents must be of a completely different nature from the perception and use of concepts that characterize us as knowers.}\]

\[\text{Schopenhauer thus argues that our nature as individuals is composed of two heterogeneous elements: that of a knowing subject, and that of a willing subject. It is feeling that characterizes the latter, and feeling is an immediate modification of the will – immediate, in that, unlike modifications of the knowing subject, feeling is not mediated by any relationship to an object. Now, underlying all modifications of will, i.e., all feelings, is one and the same motivational force: the insatiable striving that characterizes the will-in-itself. Whatever the explicit motive, every desire is ultimately an expression of this striving: ‘The motive stands everywhere before the will as a multiform Proteus: it always promises complete satisfaction, quenching of the will's thirst.’}^{3}\]

\[\text{This insatiable striving for satisfaction is, however, manifested differently in different characters. Each person has a unique character that determines which motives move her to which actions; this character is a causal law, binding motives to actions, in exactly the same sense as, e.g., the law of gravity binds gravitational causes to their effects. Schopenhauer, in fact, explains motivation as ‘causality seen from within,’ so that the feeling of desire, or striving, that binds motives to actions is simply the ‘inner’ perspective on causal necessity. Schopenhauer thus attributes will, at least by analogy, to all phenomena, conceiving the will’s striving as the very glue of the causal necessity that reigns throughout the empirical world. For animals – including humans – perception is the medium through which motives bring about action, though, for humans, reason expands this medium through the calculation of consequences not given in immediate perception. In this way, the subjection of the understanding and reason to the will’s incessant striving constitutes us as agents.}\]

\[\text{Now, Schopenhauer, like Kant, holds to a universal causal determinism throughout the empirical world. Schopenhauer’s conception of agency fits smoothly with this determinism, since he takes the character governing any individual’s actions on exactly the same model as other laws of nature. Agency, being motivated to particular actions, is the inner perspective on causal determination, in a sense simply ‘what it feels like’ to be causally determined.}\]

\[\text{Schopenhauer’s relating of rational and causal orders is here thoroughly consistent. The rational is always assimilated, or subordinated, to the causal, whether as mirror of the world in cognition, or as medium through which causality acts in motivation. In both cases, Schopenhauer interprets the activity}\]

\[^{1}\text{WWV, p.99, italics are Schopenhauer’s.}\]

\[^{2}\text{See especially WWV, §§17-18.}\]

\[^{3}\text{WWV, p.386.}\]

\[^{4}\text{VW, §43, pp.176-7: ‘motivation is causality seen from within. [...] This insight is the foundation stone of my entire metaphysics.’}\]
of the understanding\(^1\) and reason as grounded in the causal empirical realm: the understanding properly applies the concept of causality only in empirical perception, and reason operates by abstraction on such perception, and by manipulation of the resulting abstractions.

Kant, by contrast, insists on keeping the rational and empirical-causal orders carefully distinct in his theoretical philosophy. However, in his practical philosophy, Kant runs into serious difficulties when trying to maintain this distinction in coordinating the first and second Critique conceptions of freedom. For he effects this coordination through the ‘mechanistic’ conception of the inclinations considered above, and through the claim that practical reason and the moral law cannot infringe on the mechanism of the sensible world.\(^2\) Yet the role of the inclinations in Kant’s considered account of motivation in the second Critique cannot be understood in terms of mechanical push-and-pull. Moreover, as we’ve seen above, Kant’s moral psychology involves in an essential way the conflict of rational self-determination with the inclinations. While there is, for Kant, no opposition as such between moral action and satisfaction of the inclinations, he does insist that the moral disposition must acquire strength through cases of genuine conflict between duty and inclination, in which the agent develops her moral character by striking down, through pure practical reason, the influence of the inclinations.\(^3\)

But if the inclinations belong to the realm of phenomenal causality, such conflict between practical reason and the inclinations would seem to violate Kant’s doctrine of mutual non-infringement. Moreover, it would seem also to confuse the rational and empirical-causal orders in a way that would undermine any dismissal of Schopenhauer’s critique on the grounds of a similar confusion. Just as the manifold of intuition cannot, according to the Kantian position, consist of empirical-causal events if it is to be a manifold in which the understanding acts, so the same would seem to be true of the inclinations, i.e., that they cannot simply belong to the realm of phenomenal causality, if they are the kind of thing whose motivational pull reason can ‘humiliate’ and ‘strike down.’

**Conclusion**

Kant and Schopenhauer, like many 18\(^{th}\) and early 19\(^{th}\) Century thinkers, wanted to at least partially incorporate human agency into a causally-deterministic empirical world conceived along mechanical-Newtonian lines. The simplest and perhaps most plausible way to accomplish this is to take pleasure and pain as basic motivational forces pulling and pushing us to act. Schopenhauer unambiguously takes this line, insisting that only pleasure and

\(^1\)I leave aside here the crucial, though somewhat mystical, liberation of the ‘pure knowing subject’ from the service of the will, described at the end of Book 4 of WWV.

\(^2\)KpV, 5:43.

\(^3\)KpV, 5:147.
pain can ever move the will,¹ and that our experience as willing subjects amounts to seeing causality ‘from within.’ Kant, too, grounds the motivating force of the inclinations in feelings of pleasure and displeasure, and suggests that a causally deterministic account of agency would be justified if the inclinations were our only source of motivation. He also argues, however, that practical reason can determine the will independently of, and in conflict with, the inclinations, and in doing so, he describes the inclinations and their influence on the will in decidedly non-mechanistic terms. Moreover, we can point to effects in the empirical realm, e.g., the feeling of respect for the moral law and – at least potentially – action against one’s inclinations, that necessarily refer, according to Kant, to a source of motivation and a causality independent of the empirical realm. This puts pressure on the Kantian position to reconsider either the strict distinction between rational and empirical-causal orders (whereby the former cannot produce particular effects in the latter), or the ambiguous conception of the inclinations that has them straddle this distinction. Kant would surely not agree with Schopenhauer’s resolution of these difficulties. Schopenhauer’s critique, however, at the very least raises questions about the Kantian position that are more serious than we might initially suppose.

Bibliography

Kant’s works cited:


Schopenhauer’s works cited:


¹FM, §16, p.231.
CHAPTER FOURTEEN

The Logic of Heidegger’s Rectorship Address

Richard A. Hearn

Immediately following his resignation as rector at the University of Freiburg in the spring of 1934, Heidegger taught a course on “logic.”¹ Just a year earlier when Heidegger assumed the rectorship of the University of Freiburg in May of 1933, he delivered his Rectorship Address entitled “The Self-Assertion of the German University.”² Heidegger’s 1934 logic course appears to continue the development of the Greek understanding of “logos” originally developed by Heidegger during his engagement with Aristotle’s Metaphysics Θ 1-3 during the summer of 1931.³

By analyzing Heidegger’s Rectorship Address in the context of his 1931 course on Aristotle’s Metaphysics and his 1934 course on logic, we may gain a better understanding of the relationship between Heidegger’s philosophy and his politics as expressed in that address.

Language Prior to the Rectorship Address

*Logos* plays an essential role in Aristotle’s analysis of *dunamis* translated by Heidegger as “force” (*Kraft*). At the beginning of his discussion of Aristotle’s Metaphysics Θ 2, Heidegger proclaims that because “the forces belong to and co-constitute the being and the being character of the realms of being that are encountered . . . the division of forces can be accomplished by following the division of realms of being.” (AM, 99). The two divisions of being discussed are (1) the division between those beings with and without discourse (*logos*-*alogos*) and (2) the division between those beings with and without souls. (AM, 99-100) Heidegger then sets out to distinguish between *dunamis logos* and *dunamis alogos*.

*Legein*: to glean, to harvest, to gather, to add one to the other, and so

to place the one in relation to the other, and thus to posit this relationship itself. *Logos*: the relations, the relationship. . . *Logos* means therefore, rule, law, yet not as something which is suspended somewhere above what is ruled, but rather as that which is itself the relationship: the inner jointure and order of the being which stands in relation. *Logos* is the *ruling structure*, the *gathering* of those beings related among themselves. (AM, 103)

Here Heidegger claims that *logos* acts as a force gathering beings into the realm of being in which the ruling structure for the realm will be the same *logos* originally responsible for the realm’s gathering. But, the beings discoverable within any such realm are what they are only as gathered in that realm. Prior to being gathering by this force (*dunamis*), neither the realm nor the beings gathered within it are actual beings.

Heidegger then contends that *logos* originally had two meanings and its meaning as “relation (unified, gathering, coherence, rule) [was] therefore ‘prior’ to its meaning as discourse”. (AM, 3). *Logos* should therefore not be understood as “‘reason,’ ‘judgment’ and ‘sense’ [because they all miss its] “decisive meaning: gathering, joining and making known.” (AM, 103-104)

Heidegger asks whether the overlooked, but, more original meaning of *logos* as “gathering joining” was immediately accompanied in the history of the word’s origin by the commonly understood primary meaning of *logos* as “gathering saying”. (AM, 104) Heidegger expands the meaning of “discourse” by claiming that “[l]ogos is thus discourse in the utterly broad sense of the manifold making known and giving notice [Kundgeben]—“*conversance*” [Kundschaft]. (AM, 103) In doing so, Heidegger understands the force at play in such discourse as being ontologically prior to the human being who, like any other being, is acted upon by that force.

The other realm of beings distinguished by Aristotle as *alogos* is “[w]ithout conversance: without the possibility of taking notice, of perceiving, or of giving notice, and hence all the more not being in a position to explore and be conversant in a matter.” (AM, 104) Nevertheless, the animal without conversance does relate to its environment in a meaningful way. The hungry animal does not need language to comport itself toward its prey. The possibility of relating to the environment is not confined to the animal that has *logos*. As a particular way of relating to the environment, *logos* can be understood as ruling not only what human beings say, but, more originally understood as a “gathering joining” ruling what there is to be said by human beings.

To better understand the *logos* proper to man, we should note Heidegger’s description of the human (besouled) being that has *logos*.

When we speak of the besouled being who has *logos*, we do not mean that *logos*, conversance (discourse), is merely added on; rather, this *echein*, having, has the meaning of being. It means that humans conduct themselves, carry themselves, and comport themselves in the way they do on the basis of this having. This
The Logic of Heidegger’s Rectorship Address

*echein* means having in the sense of governing over . . . to be empowered for conversance and above all through conversance *(logos)* means to be conversant in oneself and from out of oneself. (AM, 108-9)

The essential difference between the animal without conversance and the human with conversance is not that conversance has been added on to the former to create the latter. Rather, the essential difference is that the human being necessarily finds itself among the beings subject to the “gathering joining” of *logos* while the animal necessarily finds itself comporting itself toward those beings governed by that same “gathering joining” from the outside.

Everything that we know, even our “possibilities and necessities”, all belongs to *logos* as conversance.

[C]hoice and decision . . . belong to *logos* as conversance, on the basis of which human beings are aware of things and investigate them. At the same time, they are aware of their own possibilities and necessities. Whenever this conversance addresses itself to things and discusses them, it is a conversance which deliberates with both itself and others . . . It is an “I” saying . . . “Language” is understood here in the broadest sense of *logos*, as a conversant gathering, as a gatherness of beings in “one”, in Dasein, which is at the same time a dissemination. (AM, 109)

In gathering us as Dasein into the world, language commands us. “[T]he human being is . . . the living being who lives in such a way that his life, as a way to be, is defined in an originary way by the command of language.” (AM, 109) Animals without *logos* live in such a way that their life is commanded by the command of nature.

But, Heidegger claims that language not only commands, but language empowers.

The human being “has the word”; it is the way he makes known to himself his being, and the way in which he sees himself placed in the midst of beings as a whole. To be empowered with language--; language, however, not merely as a means of asserting and communicating, which indeed it also is, but language as the wherein the openness and conversance of world first of all bursts forth and is. (AM, 109)

Heidegger can be understood here as claiming that, -- although language as “commanding” does limit us, -- language as “empowering” is what opens us up to the world wherein we first gain the capability to make known *(Kundmachen)* things in that world through our enactment of “conversant forces” *(Kundige Kraft)* (AM, 111) The kind of conversance *(Kundschaft)* in which things in the
world opened up by language are made known “solely for the sake of investigating them and being knowledgeable about what and how they are . . . is science; *episteme poietike.*” (AM, 111). Although all animals are empowered by nature to hunt, fly, swim, reproduce, etc., only the animal with *logos* has the capability of opening up the world by making things including itself known through conversance.

In an important section entitled “[c]apability necessarily has a realm (Bereich) and contraries that are in that realm,” Heidegger discusses Aristotle’s text distinguishing the way a conversant force empowers doctors to produce health from the way a force without conversance empowers a warm body to produce heat. Heidegger quotes Aristotle:

> And indeed forces which are in themselves conversant are always, as the same, directed at contraries. However, those without conversance are, as one, directed at a singular. For example, the warm is directed only at making warm, but the art of doctoring is directed at sickness and health. (AM, 112)

Heidegger’s point is that Aristotle recognized the essential role *logos* has in opening realms containing “contraries”. In Aristotle's example, “the art of doctoring as the healing of sickness is in itself and in fact necessarily oriented toward health.” (AM, 114) Both “sickness” and “health” as contraries are found in one and the same realm only because the art of doctoring is a conversant force. “[B]ecause the openness of the realm of force happens for this force in and through the *logos* which belongs to it, the open realm is not only completely wider [than the force without conversance], but within this realm the contrary is necessarily posited, in the relational realm of force.” (AM, 114).

The giving off of heat by a warm body occurs as a force without conversance and, for this reason, is unrelated to the overcoming of coldness. Coldness as the contrary to warmness is not removed by the warmth giving power of a warm body in the realm of nature where forces without conversance compete. Because nature is without *logos*, the force of coldness and the force of warmness are given in separate realms of being where each realm necessarily remains closed. In the case of healing in the realm opened up by the positing of health and sickness, humans and animals can also “heal” on the basis of nature “understood” as a force without conversance, *i.e.*, without contraries.

The positing of the contraries by *logos* is what originally opens up the realm wherein the human soul happens as striving. Heidegger understands *dunamis* (*Kraft*) for Aristotle to mean “that on behalf of which a transposition (*metabole*) ensues, and in such a way that the from-out-of-which (*arche*) is in another being than what is transposed.” (AM, 72). *Dunamis* is the origin of change which origin of change is in another being than the changing being itself, or, in the case where the originating being and the changing being are the same, then each is what it respectively is as a different being.” (AM, 67). In the
doctor example, *dunamis* is “that on behalf of which” the patient is healed because “the from-out-of-which (*arche*)” of that healing is found in another, the doctor. In this example, that on behalf of which the patient is transposed from sickness to health, the ground of the change in the patient brought about by the doctor, is a *dunamis logos*. But, striving for health does not presuppose a *dunamis logos* in which health is posited as the contrary of health. Although animals can never be “sick”, they do “heal” in the sense that they can regain some lost natural capacity without *logos*.

Aristotle understands humans and animals to have souls. Soul “constitutes the being of beings that have the character of living” where “living” is understood here as the “fundamental character of self-moving.” (AM, 128) That which the soul strives after in self-moving, the *orekton*, “is what properly does the moving; *it is the arche of kinesis that the soul has.*” The soul’s having (*echein*) is not a property of the soul “but, having something in the manner of holding-itself-in-relation-to, of a comportment—whereby that at which the comportment is directed is made known somehow in and through the comportment itself.” (AM, 129) Although both the human and animal make known that at which their comportment is directed toward, healing from an injury, for example, only the human makes it known in relation to both that in which it is directed toward and that in which it is directed away in its being moved between the posited contraries.

**Language Immedially Following the Rectorship Address**

In his 1934 logic lectures Heidegger will claim that temporality and language necessarily happen together.

Language is the ruling of the world-forming and preserving center of the historical *Dasein* of the *Volk*. Only where temporality temporalizes itself, does language happen; only where language happens, does temporality temporalize itself. (Logic, 140)

In 1934, Heidegger makes explicit what he left implicit in his 1931 course on Aristotle: The openness of any realm of force with conversance (*dunamis logos*) happens only in the positing of contraries related *temporally*. Language therefore is “the wherein the openness and conversance of world first of all bursts forth and is” grounded in temporality. (AM, 109)

Temporality “temporalizes itself” in the opening of any realm of force with conversance providing the ground for the relatedness between the posited contraries in that realm. Temporality makes possible movement away from the contrary of sickness toward the contrary of health in the realm opened up by a particular force with conversance. In any open realm, “[l]ogos is the ruling structure, the gathering of those beings related among themselves.” (AM, 103) What Heidegger makes clear in 1934 is the primordial relationship between time understood as history and that ruling structure we have come to know as
“logic” but the ancient Greeks knew as “logos”. Heidegger begins his lecture course on logic by setting forth how the term “logic” should be understood.

Logic is, therefore, for us, not a drill for a better or worse method of thought, but the questioning pacing off abysses of being, not the dried up collection of eternal laws of thought, but the place of the worthiness of question of the human being, his greatness. (Logic, 8)

Such “pacing off” provides a “ruling structure” for any open realm of force grounded in temporality where the pace of human life can manifest itself in time surrounded by closed realms of force without conversance grounded in nothing, i.e., abyss (Abgrund). The animal lives outside the human realm looking in.

In this lecture course on logic, Heidegger distinguishes three different kinds of temporal movement associated with “history” as commonly understood: (1) flow (Ablauf, ablaufen) of the Earth, (2) process (Verfahren, Vorgang) of Life and (3) happening (Geschehen) of human beings. (Logic, 75) Heidegger claims that we speak of “the mere flow of a mechanical happening; then regarding a movement within the sphere of life – of process; [but] we speak of happening [Geschehen] in the realm of human history (Geschichte).” (Logic, 75) Because the movement of the human being within the realm of history “happens as one that is deliberate and [as one] that knows, . . . it enters into a particular lore [Kunde] of itself and is therefore, capable of being explored [erkundbar] and should be announced [künden] to others.” (Logic, 75) Heidegger claims that “[b]y lore of history, we understand the respective manner of the manifestness, in which an era stands in history, in such a way, to be sure, that this manifestness also carries and leads the historical era. (Logic, 78) Heidegger is applying his understanding of logos derived from Aristotle to the realm of human history.

In 1934, Heidegger clearly believed that the realm of human history is “the” realm underlying our understanding of who we, as a Volk (people), are in our essence. Heidegger’s goal in this lecture course immediately following his resignation as University rector is to show how this self-understanding as “historical” relates to language. He claims that “[w]e as Dasein submit ourselves [fugen uns] in a peculiar manner into the membership of the Volk, we stand in the being of the Volk, we are this Volk itself.” (Logic, 50) And what is this “peculiar” manner of submitting ourselves to the Volk? Heidegger answers by asserting that we have “completely unawares” already “submitted ourselves [fugen uns] to the moment” whenever we “pronounce ourselves, that is speak with one another” because, in such speaking, “we are fitted [eingefugrt] in these demands of the University, we will the will of the State, which itself wills to be nothing else than the sovereign will of the government and the form of government of a Volk over itself.” (Logic, 50)

To submit to something is to be resolved to it. Being “resolved” [entschlossen] for Heidegger is fundamentally different as ontologically prior to being “decided” [entscheiden]. (Logic, 66)
We are resolved to something—in this lies the fact that that to which we are resolved stands constantly before us, determining all of our being; it does not occupy us occasionally, but the resoluteness gives our being a definitive form and constancy. It is not meant with this any condition that one carries around with oneself, just as we say: He is a human being capable of deciding. In resoluteness, the human being is rather engaged (eingerückt) in the happening that is forthcoming. (Logic, 66)

“[T]he execution of an affirmative or negative choice of the today . . . is a closing of one’s mind vis-à-vis the happening, instead of an opening up of the happening.” (Logic, 66) Why is this so? Because who we are for Heidegger has been determined by our prior engagement in “the happening that is forthcoming,” that which is forthcoming can only appear susceptible to “choice” if we have closed our “mind vis-à-vis the happening.” (Logic, 66)

Everything found in a historical realm is, as historical, subject to the ruling structure of the force with conversance understood as history. But, exactly how does this or any other force exercise power over its essential realm? Heidegger answers that force with conversance reigns by controlling through that conversance all that manifests in the realm.

That in the midst of which history happens is manifest as such through the happening [of history]. It oppresses and threatens, hampers and discloses as that which is [als Seiende]. In other words, the happening is in itself lore [Kunde]—it announces [kundet] beings [das Seiende], in which it—dispersed in them—remained fitted. . . . Lore is not pasted on history from the outside, but happening as exposed-transported [ausgesetz-entruckt] is in the manner of announcing, namely that wherein history is exposed [ausgesetz], whither it is transported [entruckt]. And with this, lore . . . announces the entire happening and the situation of its moment respectively. (Logic, 132)

History manifests itself as happening in the temporal presencing of beings. But, because all presencing happens only as unconcealing, that which is forthcoming in history necessarily remains concealed until the time of its presencing as the happening of history. History is this happening, but, in its essence, history is necessarily more than ever happens.

[T]he genuine lore [Kunde] of history announces [kundet] precisely as it sets us before that which is concealed. The mystery of the moment is the lore of that which is overpowering and inevitable. (Logic, 132)

The fact that the “overpowering and inevitable” is never fully unconcealed
in history differentiates Heidegger’s understanding of time as happening (*Geschehen*) from his understanding of time as process (*Vorgang*). Without the mystery of the moment, human life would be nothing more than a process.

**Language in Heidegger’s Rectorship Address**

In his 1933 Rectoral Address, Heidegger called upon the German people to “submit to the power [*Macht*] of the beginning . . . in Greek philosophy.” (RA, 110) In that beginning, ”Western humans rise up for the first time by the power of that people’s language against *being as a whole* [das Seiende im Ganzen] and question and comprehend it as the being [das Seiende] that it is.”¹ Prior to this philosophical uprising empowered by language -- an uprising Aeschylus had already written was destined to fail [versagt] -- “being as a whole” had never been questioned by man. (RA, 110) Such questioning was the province of philosophy understood as first science. “Science is the questioning that stands firm in the midst of *being as a whole* [Seienden im Ganzen] as it continually conceals itself.” (RA, 111) But, Heidegger claims that “[i]f we will the essence of science in the sense of questioning, unsheltered standing-firm in the midst of the uncertainty of *being as a whole*, then this will to the essence will create for our people the world of its innermost and extreme danger, *i.e.*, it’s truly spiritual world.” (RA, 112) The questioning of “being as a whole” is dangerous because, when any particular “being as a whole” is questioned, that “being as a whole” works itself back into concealment.

Heidegger had claimed in his 1931 course on Aristotle and again in his 1934 course on logic that “the power of that people’s language” was the power of *logos* to open a realm wherein a people happen in the midst of beings. History understood as the realm where being happens temporally is the realm where man’s uprising against being as a whole is carried out. Heidegger claimed in his 1931 course on Aristotle that the primordial meaning of *logos* for the Greeks was as a ruling structure.

*Logos* means therefore, rule, law, . . as that which is itself the relationship: the inner jointure and order of the being which stands in relation. *Logos* is the *ruling structure*, the *gathering* of those beings related among themselves. (AM, 103)

Heidegger’s understanding of *logos* as both commanding and empowering a people, allows Heidegger to refer to the world in which that people are rooted as a “spiritual world”.

And the spiritual world of a people is not its cultural superstructure—just as little it is its arsenal of useful knowledge and

¹The original translation has been modified to translate *das Seinde* as the compound noun “being” rather than “beings”.

152
values. Instead, it is the power of deeply preserving the forces that are rooted in a people’s soil and blood; the power to arouse most inwardly and to shake up most extensively \[Erschütterung\] our people’s existence \[Dasein\]. (RA, 112)

Heidegger here expresses the troubling belief that susceptibility to the originary forces of the beginning of Western culture is somehow linked to “a people’s soil and blood.” Unfortunately, Heidegger was politically associated at the time of his Rectorship Address with a political party that linked the characterization of people for purposes of determining their rights according to “soil and blood”. The evil inflicted by that party and its followers on those peoples murdered and disenfranchised based upon “soil and blood” must be remembered in any interpretation of Heidegger’s writings.

Yet, a less troubling, although not completely benign, interpretation of Heidegger’s reference to a people’s linkage to “soil and blood” may be consistent with Heidegger’s discussion of resoluteness in the logic course.

We are resolved to something—in this lies the fact that that to which we are resolved stands constantly before us, determining all of our being; it does not occupy us occasionally, but the resoluteness gives our being a definitive form and constancy. (Logic, 66)

Heidegger is asserting that people do not “decide” what ruling logos to be resolved to, but, exist as a people only as either resolved or not resolved to the logos which “gives [that people’s] being form and constancy.” Contrary to modern thought, Heidegger is claiming that logos is always the logos of a particular people and never the logos of all people. So understood, Heidegger’s reference to “soil and blood” may be understood as claiming that the empirically observable division of human beings based upon “soil and blood” is itself grounded in the naturally occurring divisions found within logos. Heidegger’s claim would appear to be that a people achieve its “form and constancy” as the people that they authentically are and will be only so long as the people resolutely submit to the governing logos of their beginning as manifested in their “soil and blood.”

The power of any people for Heidegger is found in “the power of deeply preserving the forces that are rooted in [that] people’s soil and blood.” (RA, 112) But, how are those determining forces for a people disclosed so that the people can preserve those forces as guiding them throughout their history? In his Rectorship Address Heidegger claims that, because science was the way of unconcealing “being as a whole” in the beginning of the Western tradition, science must continue its essential role of disclosing “being as a whole” if the German people are to remain in that realm opened in its beginning. “All science is philosophy whether it knows it and wills it or not.” (RA, 110)

For the ancient Greeks, the kind of discourse in which things in the world opened up by language were made known “solely for the sake of investigating them and being knowledgeable about what and how they are . . . is science;
"episteun poietike." (AM, 111). Because such beings in the realm of the West, like beings in any realm opened by *logos*, "continually conceal themselves," the questioning unconcealment of beings in that realm is an "innermost necessity" for the people of the West. "Only if and when we resolutely obey this decree in order to win back the greatness of the beginning will science become the innermost necessity of our existence." (RA, 111)

But, Heidegger claims that a people must recognize the "impotence" of the "active perseverance" of science in the face of fate. (RA, 111) Heidegger argues that, although science is "the inner necessity of our existence," science must recognize its impotence "in the face of destiny." Science is impotent to affect the eventual outcome of Western mankind's uprising against being as a whole. (RA, 110) The uprising mounted by the Greeks which marked the beginning of Western humanity and continues as its governing *logos* today was destined from the beginning to fail. In the Rectorship Address, Heidegger cites the ancient tragedian Aeschylus as support for his conclusion: "But knowledge is far less powerful than necessity" (RA, 110)

That means: All knowledge of things remains beforehand at the mercy of overpowering fate and fails before it.

For precisely this reason, knowledge must develop its highest defiance, for which alone the entire might of the concealedness of beings will first rise up, in order [for that knowledge] really to fail. Thus beings reveal themselves in their unfathomable inalterability and confer their truth on knowledge. This adage about the creative impotence of knowledge is a saying of the Greeks . . . (RA, 110)

Knowledge fails to alter the course of a people’s destiny as determined from the beginning by the force of history for that particular people opened up by *logos*. Instead, knowledge only facilitates the coming of that destiny. In May of 1933 when Heidegger gave his Rectorship Address, Heidegger’s understanding of the West's destiny was apocalyptic.

At the end of the Rectorship Address, Heidegger claimed that, because the destiny of the West is inalterable, nobody will care whether the German people willed that destiny or not. Instead, what is important is that the German people will themselves regardless of the consequences.

But neither will anyone ask us whether we will it or not when the spiritual strength of the West fails [in its uprising against being as a whole] and the West starts to come apart at the seams—when this expended, illusory culture collapses into itself, pulling all forces into confusion and allowing them to suffocate in madness. Whether such a thing does or does not occur depends solely on whether we as a historical-spiritual people will ourselves, still and again, or we will ourselves no longer. (RA, 116)

Heidegger’s claim in the Rectorship Address therefore is not that the
German people could somehow help the West avert its destiny. Instead, Heidegger can be understood as encouraging the German people to facilitate the coming collapse of the West by willing themselves, “still and again”, as a people.

This understanding can explain Heidegger’s reference to the “glory and greatness of the “new” beginning found in the concluding paragraph of his Rectorship Address.

We can only fully understand the glory and greatness of this new beginning, however, if we carry within ourselves that deep and broad thought upon which the ancient wisdom of the Greeks drew in uttering the words: “All that is great stands in the storm . . . (Plato, Republic 497d9) (RA, 116)

Because new beginnings happen only in or as the collapse of previous beginnings, any new beginning for the West must await the collapse of existing Western culture into itself, “pulling all forces into confusion and allowing them to suffocate in madness.” (RA, 116) Such a happening can only be brought about by philosophy’s questioning of being as a whole. Therefore, if Heidegger believed that Hitler’s rise to power was relevant to the destiny of the West, that relevance would have been limited to Hitler’s having allowed the West “to suffocate in madness” as the necessary precondition for a “new beginning.” (RA, 116) While Hitler may have believed his rise to power heralded a new beginning to come, references in the Rectorship Address to the inability of knowledge to alter a people’s destiny suggests that Heidegger may have understood the rise of the Third Reich to be but another step toward the destined collapse of the West. But, because even Heidegger’s “knowledge of things remain[ed] beforehand at the mercy of overpowering fate,” Heidegger lived to see neither the collapse he had predicted nor the “new beginning” he had hoped for. (RA, 110)
On the frontier between two agencies, where the first passes over to the second, there is a censorship which only allows what is agreeable to it to pass through and holds back everything else….what is rejected is in a state of repression. --- Sigmund Freud

In Sartre’s critique of Freud’s censor he argues that the very idea of a censor is problematic, because it requires that this agency know and not know at the same time. In this paper I review how Freud developed the notion of the censor and its role in his psychoanalytic theory. I then discuss Sartre’s critique of the censor as both an impossible psychoanalytic fiction and as an exercise in bad faith. I proceed to show how despite Sartre’s convincing critique, Lonergan incorporates Freud’s censor into his account of levels of consciousness. While both Freud and Sartre consider the censor to be primarily repressive, Lonergan understands it to have a dual role, constructive as well as repressive. I elucidate the constructive role of the censor by noting its similarity to Aquinas idea of a sensory power, the cogitative. I argue that by seriously considering the constructive role of the censor, the epistemological problem of the censor can be resolved.

The Problem of the Censor

Sigmund Freud first used the metaphor of a censor to designate a repressive and filtering function of the psyche in his pivotal work *The Interpretation of Dreams* (1899). The censor became a technical term of psychoanalysis and a familiar term in popular culture in the early Twentieth Century. Jean-Paul Sartre in *Being and Nothingness* (1943) leveled a blistering attack against Freud’s concept of an independent psychic agency operating within consciousness. Yet, ten years later Bernard Lonergan in his work *Insight: A Study of Human Understanding* (1957) integrated the concept of the censor into his account of psychic consciousness. This raises the question of whether Lonergan’s continued use of Freud’s concept was simply the result of a critical lapse or whether he found something about the censor that warranted its continued inclusion in a rigorous account of the psyche. As I will show, in Lonergan’s nuanced account of conscious intentionality as differentiated, as multi-leveled, there is a psychic level of consciousness with its own functions
and operations. The censor is not seen, as it is for Sartre, as an interloper in an otherwise homogeneous intentionality. In addition, Lonergan developed a view of the censor as having a dual function, constructive as well as repressive. The constructive role of the censor, which neither Freud nor Sartre describe, has barely been studied since Lonergan articulated it over fifty years ago.

Before we examine what Freud, Sartre, and Lonergan mean by the censor, let us consider in a preliminary way an inherent difficulty with the very concept of psychic censorship. As repressive, the censor blocks from consciousness the affect-laden images which would lead one to a disturbing or painful realization. How is it possible for the censor in its repressive role to block from consciousness the images that would give rise to an unwanted insight, if one does not already have some apprehension of the insight which one then does not want? This question, incidentally, is the converse of the sophistical challenge in the Meno regarding the impossibility of arriving at knowledge:

How will you look for something when you don’t in the least know what it is...how will you know that the thing you have found is the thing you didn’t know? (80d).

The problem with the censor is how to account for the proleptic apprehension of an understood which the censor blocks us from understanding. We can highlight the difficulty more sharply by reaffirming with Lonergan the unity of the developing, existing subject.¹ The higher levels of conscious intentionality on which I question and understand are not more ‘me’ than the underlying psychic level:

Nor are the pure desire [to know] and the sensitive psyche two things, one of them ‘I’ and the other ‘It.’ They are the unfolding on different levels of a single, individual unity, identity, whole. Both are ‘I,’ and neither is merely ‘It.’ ²

The censor cannot be conceived of as a psychic function, distinct from me, which stands guard at the gate separating me from the rest of the unconscious. The I who desires to know and to understand is the I who interferes with the images that would enable me to understand. This unity which I am seems both to desire to know and to flee understanding. That the one same subject would have conflicting desires is not in itself a problem. It is a manifestation of the fact that the subject is, according to Lonergan, a unity in tension.³ The difficulty is that the concept of the censor requires that one would apprehend and understand and would neither apprehend nor understand the same thing at the same time.

In his The Interpretation of Dreams, Freud primarily developed a theory of

¹Lonergan, Insight, p. 495.
²Lonergan, Insight, p. 499.
³Lonergan, Insight, pp. 494-504.
dreams and secondarily suggested a general theory of the mind or psyche. The key insight into dreams revealed through the laborious work of their interpretation is that “a dream is a fulfillment of a wish.” The concept of the censor emerges in Freud’s account of how repression operates in dreaming and in certain other psycho-pathological structures. In his summary of the psychology of repression in *On Dreams*, Freud describes two “thought-constructing agencies”—one has free access to consciousness and its contents, the other is in itself unconscious and can only reach consciousness through the former consciousness. The censor functions on the borderline between these two agencies:

On the frontier between the two agencies, where the first passes over to the second, there is a censorship which only allows what is agreeable to it to pass through and holds back everything else. According to our definition, then, what is rejected by the censorship is in a state of repression.\(^1\)

In dreams the strict control of the censorship, which Freud likened to the Russian censorship of his day,\(^2\) is somewhat relaxed allowing certain impulses and desires to be expressed in dream representations. This enables contents of the unconscious a circuitous way into consciousness. Yet, while the censorship is relaxed in sleep, it is not completely eliminated. Thus, dreaming involves a systematic process of distortion and disguise which renders the content of a dream obscure. For example, a dream content can be disguised by the use of its opposite. The fundamental pattern for the generation of dreams is “repression—relaxation of the censorship—the formation of a compromise.”\(^3\)

The intentions of one agency (the conscious mind) and the demands of the other agency (the unconscious mind) arrive at a compromise in the obscure dream representations. One vital function of the disguise of the dream contents is to enable the dreamer to remain asleep. When the dreamer awakens the censor regains its strength and can eliminate any trace of the partially revealed, previously unconscious contents. Freud surmises that this is one explanation for the complete or partial forgetting of dreams.

Freud readily acknowledges that the psychology of repression that he outlines is but a crude hypothesis based on a simile:

We have gathered an impression that the formation of obscure dreams occurs *as though* one person who was dependent upon a second person had to make a remark which was bound to be disagreeable in the ears of this second one; and it is on the basis of this simile that we arrived at the concepts of dream-distortion and

---

censorship.¹

Yet, he is confident that further research into the two opposing psychic agencies will ultimately confirm it. Of one thing he is sure, that “the second agency controls access to consciousness and can bar the first agency from such access.”² The second agency, as was indicated above, is that which has free access to consciousness and its contents, the first agency is wholly unconscious. In attributing the control of the access to consciousness to the second agency, he is characterizing the censor as a function of consciousness. But in what sense is the censor conscious? Am I responsible for the censorship; must I be continuously vigilant to guard the gates of consciousness? To determine in what sense the censor is conscious, let us examine what Freud means by the unconscious.

In the course of systematizing psychoanalysis Freud developed two typologies of the psyche. His accounts of the nature and role of the unconscious and the censor evolve with the revisions to his metapsychology. First, it should be noted that the unconscious has two senses: a descriptive sense in which simple awareness is its criterion; and a systematic sense, the main interest of psychoanalysis, in which the unconscious is seen as the source of motivation and psychic conflict.³ The latter is the unconscious as uncovered through the clinical phenomena of resistance and transference, and as manifested in dreams and in certain waking parapraxes, such as slips of the tongue, misreadings, slips of the pen (typos), bungled actions, etc.⁴

The first typology was initially one of regions or systems: the unconscious, the preconscious, and the conscious. In The Interpretation of Dreams, dream thoughts, for example, are described as in the unconscious. The censor, as we have seen, is placed at the border between the unconscious and the conscious. In his later essays in metapsychology, the typology of static places is gradually replaced by a dynamic economics of psychic forces. In The Unconscious (1915), Freud differentiates the Preconscious from the Unconscious and the Conscious. We find that the censor plays a pivotal role in the interrelation of these psychic systems. An unconscious psychical content is first tested by the censor. If rejected, it is repressed and remains unconscious. If it passes the censor’s testing, it enters the system of consciousness and though not yet fully conscious, it may become conscious. This probationary state within the conscious system is called the preconscious. The censor plays a role in the initial screening and repressing of contents, and then in determining whether what is preconscious becomes fully conscious.⁵ The Unconscious in the systematic sense is not merely what we would describe as unconscious. It is differentiated into that which is initially or temporarily unconscious and may become conscious and that which is repressed and cannot become conscious.

¹Ibid
²Ibid.
Similarly, the Conscious in the systematic sense includes what we might describe as conscious, what we are aware of, as well as that which has been repressed, which we are not presently aware of, but yet still may become conscious if not again repressed (the Preconscious). It is clear that these three psychic systems are differentiated not so much by our awareness of the acts or contents, but by the process of repression and the agency of repression, the censor.

As Freud’s psychology expanded to include a theory of society and culture, he developed a typology that was a kind of ‘personology.’\(^1\) It was not conceived of in terms of regions but rather in terms of roles of the individual. In his later works rather than divide the psyche into the Unconscious, the Preconscious, and the Conscious he now develops the familiar triad of the Ego, the Id, and the Superego. Freud did not abandon the earlier elements of the first typology, but he saw them as less critical in his definitive statement of the structure of the psyche. There is not a simple correlation between the respective elements of the two typologies. The Ego is not to be identified with the conscious because it starts out as preconscious and the preconscious remains its nucleus.\(^2\) Further, the Ego is in part unconscious because it is not sharply separated from the Id. The lower portion of the Ego, the basic passions, merges into the Id. Not only are the nether reaches of the Ego unconscious, even the highest intellectual and critical operations of the Ego can be unconscious as, for example, when a difficult intellectual problem is resolved during sleep. Similarly, while the Id and the Superego are generally considered to be unconscious, the conscious extends into the Id insofar as the repressed becomes a part of the Id: “The repressed is only cut off sharply from the ego by the resistances of repression; it can communicate with the ego through the id.”\(^3\)

Notwithstanding, the reciprocal contamination of the unconscious Id and the conscious Ego, Freud reasserts that the division of the psychical into what is conscious and what is unconscious is fundamental to psychoanalysis. How does the agent of repression, the censor, function in the context of the second typology? In each individual there is found “a coherent organization of mental processes” which Freud defines as the Ego. He further characterizes the Ego as that to which consciousness is attached; as the agency in control of the individual’s operations with the external world; and as the mental agency capable of supervising all of its conscious mental activity. In this last role of the self-conscious agency in control of its own mental operations and contents, the Ego is the source of all repressions. “Even [when it sleeps] it exercises the censorship on dreams.”\(^4\)

Reviewing the two typologies of Freud’s metapsychology has brought us closer to the nature of the censor in his thought. In the first typology, the censor stands as the guard or tester on the border between the Unconscious and the Conscious, and also on the border between the Preconscious and the

---

\(^1\)Ricoeur, p. 181.
\(^3\)Ibid. p. 635.
\(^4\)Ibid. p. 630.
Conscious. The censor is a function of the second thought-constructing agency, that which has access to consciousness. In the second typology, the role of the censor is attributed to the Ego, to which consciousness is primarily attached, but which at its outer edges is also unconscious. The function of the censor, then, is closely associated with consciousness in both typologies, that is, throughout the development of Freud’s metapsychology. Yet, whether the process of censorship is consciously controlled by the self-conscious ego remains unresolved. While guarding and protecting consciousness and functioning as an agent of consciousness in that sense, it nevertheless seems that censorship itself takes place behind the back of the conscious ego.

The critique of the psychoanalytic hypothesis of the censor found in Sartre’s *Being and Nothingness* is an integral part of his denial of the unconscious. His denial of the unconscious is one of the most striking aspects of Sartre’s philosophy of consciousness especially in light of the fact that he proceeds towards the end of the same work to develop his own existential psychoanalytic method. Without examining fully what Sartre means by ‘consciousness’ we can at least recall that he distinguishes between prereflective and reflective consciousness. It should be noted that these are not to be understood as distinct, co-existing regions of the human psyche. Consciousness is unitary and homogeneous, what differentiates the prereflective from the reflective is the object of the intentionality of consciousness.1 While prereflective consciousness and reflective consciousness are not to be understood as parts of consciousness, the nature of reflective consciousness does introduce the possibility of self-deception.

As consciousness turns on itself in reflection, a gap is introduced into the unity of consciousness—a duality is introduced in consciousness. The duality in the unity of consciousness is what makes possible the self-deception that Sartre calls ‘bad faith.’2 Reflective consciousness and the duality it introduces do not necessarily create bad faith, just the conditions for bad faith. While not absolutely ineluctable, bad faith can be so durable and long-lasting that it becomes the normal mode of existence. The fundamental goal of bad faith is: “To cause me to be what I am, in the mode of ‘not being what one is,’ or not to be what I am in the mode of ‘being what one is.’”3 Authenticity is possible, according to Sartre, but only as a recovery from the inauthenticity of being in bad faith.4

Freud’s psychoanalytic theory which posits a conscious ego and an unconscious id is, for Sartre, a theoretical exercise in bad faith, an instance of meta-psychological self-deception. Rather than contend with the self as lying to itself in reflective consciousness, Freud’s theory introduces into the psyche an intersubjective structure, which renders the analysand as much a passive

---

3Ibid. p. 110.
(and thereby guiltless) spectator as the analyst:

Thus psychoanalytic theory substitutes for the notion of bad faith, the idea of a lie without a liar; it allows me to understand how it is possible for me to be lied to without lying to myself...; it replaces the duality of the deceiver and the deceived, the essential condition of the lie, by the “id” and the “ego.”

Freud further complicates the theoretical deception by introducing a third agency—the censor. Sartre characterizes Freud’s censor as “a line of demarcation with customs, passport division, currency-control, etc.” If the censor is in control of the distortion and obfuscation in my consciousness, then I need not realize nor take responsibility for my own self-deception. However, the problem of bad faith is not thereby eluded, because “in order to overcome bad faith, it has established between the unconscious and the conscious an autonomous consciousness in bad faith.”

The censor functions as a second duality placed between the duality of the ego and the id. The id is seen by Freud, according to Sartre, as a blind libidinal force, and the ego, specifically the analysand, as a passive spectator of the psychic phenomena brought to light. The agency responsible for the deception now becomes the censor. The same duality that must operate in self-deception, the deceiver and the deceived, is relocated in condensed form in the censor (a kind of autonomous ‘mini-me’). In order for the censor to effectively repress, it must know what it is repressing:

How could the censor discern the impulses needing to be repressed without being conscious of discerning them? How can we conceive of a knowledge which is ignorant of itself?

In fact, the censor must not only be aware of that which it is repressing, it must also choose what to repress. It allows acceptable desires and impulses, even sexual, into consciousness, and incarcerates those that are unlawful.

Freud’s censor, then, for Sartre does nothing to alleviate the philosophic and moral problem of bad faith. The philosophic difficulty of how one can know and not know the same thing at the same time is not resolved but displaced. Similarly, the moral problem of taking responsibility for one’s own desires, impulses and feelings is passed off to another, a fictitious conscious agent who chooses for me. The very hypothesis of the censor, for Sartre, is an exercise in bad faith.

In light of Sartre’s seemingly devastating critique of Freud’s censor, we might question why Lonergan would employ this conception in his major work *Insight*. His acceptance of the essential meaning of the censor becomes

---

1Ibid. p. 92.
2Ibid. p. 90.
3Ibid. p. 94.
4Ibid. p. 93.
understandable, if we consider two significant differences in Lonergan’s and Sartre’s interpretations. First, Lonergan does not deny the existence of the unconscious, but he does not mean by the unconscious a region of the psyche (the human mind), but rather an underlying neural level. The fact that Lonergan acknowledges the reality of brain functioning as providing materials for psychic integration manifests his acceptance of the self’s embodiment. In *Method in Theology* (1972) Lonergan articulates a second meaning of the unconscious: “the twilight of what is conscious but not objectified.” However, in *Insight* he uses the term ‘preconscious’ for that of which we are only dimly aware or only potentially aware insofar as it is receives no advertence. Lonergan’s ‘preconscious’ in *Insight* corresponds to Sartre’s ‘prerreflective consciousness.’ Both of these terms roughly correspond to Freud’s notion of ‘the preconscious’, although neither Sartre nor Lonergan conceive of the preconscious or the prerreflective as a region of the psyche coexisting with consciousness.

Secondly, Lonergan does not hypostasize the censor as an autonomous consciousness as Freud does, according to Sartre. In his multi-leveled metaphysics of the developing self, there are supervening levels each governed by proper laws and proper content, and each is the subject of a distinct science:

> In man, there is intellectual development supervening upon psychic and psychic supervening upon organic….In the organism both the underlying manifold and the higher system are unconscious. In intellectual development both the underlying manifold of sensible presentations and the higher system of insights and formulations are conscious. In psychic development the underlying neural manifold is unconscious and the supervening higher system is conscious.  

Further, on each level of integration we find a network of correlative terms defined by the laws of that level. Lonergan calls such explanatory terms ‘conjugates.’ Rather than acting as some kind of existing, autonomous agent, the censor in Lonergan’s account functions as an explanatory conjugate of the psychic level on a par with impulses, affects, images, behaviors, dreams, and the processes of repression and inhibition. That censoring is not an operation experienced by the conscious subject is not an argument against its role on the psychic level.

Lonergan appropriates the essence of Freud’s account of the censor in its repressive role as barring certain psychic materials from becoming conscious: “The function that excludes elements from emerging in consciousness is now familiar as Freud’s censor.” He also agrees with a number of auxiliary points Freud makes about the censor, for example, that it is relaxed in sleep allowing

---

1Lonergan, *Method in Theology*, p. 34 fn. 5.
3Ibid. pp. 102-105.
4Ibid. p. 214.
materials to arise in dreams that it would normally reject in waking life. However, Lonergan views the repressive role of the censor to be only half of the story. He considers the primary role of the censor to be constructive. The constructive role of the censor is a function that neither Freud nor, consequently, Sartre in his critique of Freud’s censor stress.

In order to understand Lonergan’s censor as primarily constructive and as aberrantly repressive, we should review the elements and interplay of the underlying, unconscious neural level and the supervening psychic and intellectual levels of consciousness. Lonergan’s remarks about the censor are found in his account of the psychic level of the subject in the dramatic pattern of experience. As dramatic the self finds itself in an interpersonal context in which its own living is a creative work of art. The effort to maintain oneself in this social context requires insights into oneself and one’s interpersonal relations and it can involve, as well, the avoidance of those realizations that would disrupt the status quo of one’s dramatic achievement to date.

Three interrelated levels of operation are at play in the dramatic pattern of experience. First, there is the underlying, unconscious level of neural stimuli, patterns, and processes which demand some representation and integration on the supervening level of psychic consciousness. The neural demand for psychic representation is a significant requirement of the healthy subject: “The demand functions of neural patterns and processes constitute the exigence of the organism for its conscious complement; and to violate that exigence is to invite the anguish of abnormality.” Lonergan names the neural demands for conscious expression ‘neural demand functions,’ because their need for higher integration can be met in a variety of ways, primarily, in the affects and images of the psychic level of consciousness. Imaginative representations are required for an act of understanding; insights are into images not into affects per se. The next higher level of consciousness, the level of intelligent consciousness, which is the level of questioning, having insights, and formulating those insights, supervenes and integrates the underlying psychic material that has already emerged into consciousness. One’s desire to know will penetrate the underlying psychic level and even the neural level positively to seek out the patterns and images required for insight. On the other hand, the fears and biases of the dramatic subject can thwart the emergence into consciousness of the images that would give rise to unwanted insights: “Besides the love of light, there can be a love of darkness.” In its constructive role, the censor selects and arranges materials that emerge in consciousness in such a way that an insight could occur. In its repressive role, the censor blocks the emergence into consciousness of those images which would likely give rise to insight.

The avoidance of unwanted insights can be conscious and deliberate or “prior to conscious advertence” and spontaneous. The former is the root of

---

1 Ibid. p. 218.
2 Ibid. p. 211.
3 Ibid. p. 214.
4 Ibid. p. 214.
5 Ibid., p. 216
individual bias; the latter is the root of dramatic bias. When the refusal to recognize something about oneself is conscious, it is not a refusal to understand but normally a deliberate shift of one’s focus and preoccupation. In fact, the more one consciously denies something and seeks to overlook it, the more likely it is that one will continue to advert to it. There is no interference with or lasting influence on the underlying psychic level, and consequently the imaginative representations required for the insights will readily emerge. But in the case of repressive censorship, schemes that would suggest the unwanted insights are blocked from emerging into consciousness. The resulting aberration can play havoc with the normal satisfaction of neural demand functions. On a primitive, preconscious level patterns and arrangements of psychic materials are distorted and excluded.¹ The consequent aberration of repression and inhibition interferes with one’s conscious performance in the dramatic sphere, and not uncommonly in one’s other spheres of performance, in the biological, artistic, intellectual, and mystical patterns of experience. For an example of the effect of repression on the subject as biological, we need only recall Freud’s remarkable discovery of the aetiology of very serious physical disorders, such as temporary blindness and partial paralysis. Through his various psychoanalytic methods of releasing repressed psychic materials, he was able to reverse these psychosomatic symptoms in his analysands.²

We find in Lonergan’s *Insight*, then, an account of the censor that articulates both its positive and its possible negative function:

The liberation of consciousness is founded on a control of apprehensions; as has been seen the censorship selects and arranges materials for insight, or in its aberration excludes the arrangements that would yield insight.³

Lonergan’s focus in his account of the censor is on its role in aiding or thwarting the emergence into consciousness of the images that would lead to insight. In both its positive and its negative functions, the censor for Lonergan is correlated to the level of intelligence. The censor prepares the psychic materials into which we have insights, and its selection is influenced by the desire to understand or by the dread of understanding, which penetrates to the underlying psychic and neural levels. While Freud tends to identity the ‘real’ drives of the psyche exclusively with the libidinal and instinctual drives of the Id, Lonergan acknowledges and stresses the desire to know as a fundamental drive of human existence. In his emphasis on the act of understanding and the desire to know, he parts company not only with Freud, but also with Freud’s critic, Sartre.

The views of Freud, Sartre, and Lonergan can be summarized on two

---

¹Ibid. p. 216. Lonergan actually uses the term ‘unconscious’ in this passage; however, the discussion is about the censorship and aberration that are “operative prior to conscious advertence,” which I understand to mean the preconscious.


The Problem of the Censor

issues: the relation of the censor to consciousness, and the primary function of the censor. On the first point, for Freud the censor is a psychic agency which functions in an unconscious dimension of the conscious ego. For Lonergan, it is conscious but not objectified. It is not wholly conscious in the sense of intelligent and rational, nor is it a blind unconscious neural function. It is preconscious in the sense of functioning prior to our advertence. For Sartre, all intending is conscious and self-conscious, and so the censor, if there were such an agency, would have to be conscious and self-conscious in order to select and reject materials. As we have seen, Sartre rejects the censor as de trop; its machinations can be carried out by the conscious self in bad faith. For Lonergan, the censor is an explanatory conjugate of the psychic level of consciousness. It functions in correlation to the neural demands of the underlying unconscious level and to the intelligent and rational (or obfuscating and biased) intentionality of the higher levels of intelligent and rational consciousness.

On the second point, for Freud the censor is an exclusively repressive psychic agency which functions in an unconscious dimension of the conscious ego. For Sartre the censor is a fiction. Any censoring is actually carried out prereflectively by the self in bad faith, that is, as engaged in self-deception. One’s self-censoring is primarily repressive as well as being self-deceptive. For Lonergan, the censor can be resurrected as an element of the psychic level of consciousness, but its primary function is constructive, and only as aberrant is it repressive.

In conclusion, Lonergan’s concept of the censor as primarily constructive warrants some further consideration. While a constructive censor may appear to be a novel post-Freudian idea, it is not without its philosophic precursors. Indeed, we can find similar but distinct notions in Aquinas and Kant.

The role of the vis cogitativa, in St. Thomas bears striking similarities to the constructive role of the censor. In Lonergan’s work Verbum: Word and Idea in Aquinas (1946–49), he asserts that without the cogitativa there would be no insight. When Lonergan first introduces Thomas’s concept of the cogitativa, he distinguishes the activity and the object of the cogitative sense by comparing the knowledge of the technician who knows the abstract universal, which is consequent to insight and the man of experience who knows the universal in the particular, which is not intellectual but sensory knowledge under the influence of intellect. The cogitativa is a sensory act, its object is the universal in particulars, and it is under the influence of the intellect. Aquinas distinguishes this particular sensitive faculty from the other five senses and from the imagination. This faculty is found in animals and in man. In animals it is called the estimative sense, and in man, the cogitative sense. In both animals and man, the object of this sensory power is the individual thing as existing in a common nature, that is, the universal in a particular. In animals the


2Lonergan, Verbum, p. 43-44.

3St. Thomas Aquinas, In II de Anima, lect. 13, #398.
apprehension of these sensory particulars is restricted to a sense of the object as beneficial or harmful. In man, the object of the cogitativa is not so restricted; the object of the cogitative sense is any universal in a particular.

Lonergan views the cogitativa as leading to and necessary for the act of insight. In describing human intelligence as process, as non-angelic, he writes of “the experience we all have of working from, and on, a sensible basis towards understanding.” In the process of reasoning towards an understanding, the intellect is dependent on sense for its object. In the case of coming to understand what a circle is, to employ Lonergan’s example in Insight, we rely either on a visible diagram or an imagined diagram. The act of insight into the meaning of a circle comes as an increment in a process that involves imagining spokes, rim, and hub; imagining the spokes reduced to fine lines, reduced to invisible radii. Further, the intellect relies on what he calls a “preparatory elaboration of its object.” These radii must be aligned properly: they must emanate from the center out to the perimeter; they must be of equal length; they must lie side by side, they must be so numerous as to be uncountable; etc. The cogitativa makes possible the imaginative manipulation of the radii. Prior to multiplying and moving radii in our imagination, we must apprehend each radius qua radius. Only through such apprehension and comparison can we imagine one radius to be longer and one to be shorter or two to be equal. It is the cogitative preparation of the image that prepares us for insight. In fact, according to Peghaire in his work on the cogitative sense, the preparatory role of the cogitativa is to be found in the very process of reasoning itself. In order to follow an argument, the elements of the reasoning process must be placed serially so that thought can move more easily from one to the other to arrive at a conclusion.

The cogitativa and the intellect are interrelated: the cogitativa provides the necessary elaboration of the representations upon which the intellect depends for its insights; and the intellect influences through its questions and its prior understanding the apprehension of the cogitativa. How is the cogitativa distinct from the act of insight itself? The cogitative sense apprehends the universal in the particular. Insight is not an apprehension of the universal, but the abstraction of the universal. “Abstraction,” Lonergan writes, “is the elimination by the understanding of the intellectually irrelevant because it is understood to be irrelevant.” Whereas, “knowing the universal in the particular, knowing what is common to the instances in the instances, is not abstraction at all;” it is merely the operation of the cogitativa.

The preconscious censor in its constructive capacity corresponds to Aquinas’s cogitativa as that which provides the “preparatory elaboration of its

---

1 Lonergan, Verbum, p. 45.
2 Lonergan, Insight, pp. 31-37.
3 Ibid. p. 4.
5 Lonergan, Verbum, p. 53.
6 Ibid.
[the intellect’s] object.” The preparation of representations by the cogitativa is necessary for an act of understanding and for the process of reasoning. As sensory, this operation is spontaneous and natural rather than intelligent and deliberate. For Lonergan in Insight, the materials that emerge in consciousness are already subject to the preconscious selection and patterning of the censor. This arrangement of the materials results from the dialectical interplay of underlying neural demands and conscious interests and orientations. There is some process or operation, that takes places preconsciously for Lonergan and instinctually for Thomas, which selects, patterns, synthesizes, or arranges the materials that emerge in consciousness under the influence of and for higher level operations.

We find a strikingly similar notion of a pre-intellectual preparatory function in Kant’s first Critique where he argues for the necessity of a transcendental faculty of imagination. Kant describes this synthetic work of imagination as a ‘blind but indispensable function of the soul, without which we should have no knowledge whatsoever, but of which we are scarcely ever conscious’. Kant’s account of this non-conscious faculty seems very close to Thomas’s cogitativa; however, we should note that Thomas carefully differentiates the cogitativa from the imagination.

We have discussed the inherent epistemological difficulty with the notion of the censor which Sartre clearly articulates. How is it possible for a consciousness, even as prereflective or preconscious, to know and to not know the same thing in the same regard at the same time? How can the censor select material which is to be repressed without apprehending that material? If one has already apprehended the material in the process of censoring it, in what sense is it rendered unconscious? The unity of the self, for both Sartre and Lonergan, exacerbates the difficulty. The repressing work of the censor the resulting blind spots cannot be blamed on some agency apart from ourselves.

The similarity of the censor and the cogitativa contributes to resolving the difficulty of how the same self both knows and does not know in censoring the contents of its consciousness. The object of the cogitativa is the universal in the particular. There is a primitive recognition of the same in the many instances. This, as we have seen, does not require intelligence; it is a sensitive capacity we share with animals, a primitive sense of the harmful and the beneficial. Not only must the human organism be alert to positive and negative possibilities in its “external” habitat, it must also be on guard against “internal” threats to its psychic integrity. As I can recognize in the periphery of my vision and react spontaneously to swat away a spider crawling near; so, I can swat away from conscious advertence an imaginative representation that threatens my self-image. How does one recognize the image or representation as having the potential to provide material for the insight into oneself that one does not want? The intention of the understanding is already at work on the underlying levels. The desire to know helps guide the selection of materials for questions for which it seeks answers. Fear and insecurity, on the other hand, also

---

influence the selection of materials. We avoid and block those images that would lead to unwanted insights into oneself or one’s situation. The preliminary sense that there is something of a threatening nature that could be understood creates the psychic distance to block the image prior to its emergence in consciousness. The sensory cogitative act is an apprehension of a type (universal) in the representation, and it is simultaneously the apprehension of the harmful, some felt-danger to oneself.

The primitive schematizing of psychic materials seems to be the locus of the censorship both in its constructive and its repressive capacities. This censoring does not require knowing in the full, human sense of intelligence and reason, let alone rationally self-conscious choice; it simply requires an animal apprehension of possibilities. The selection and arrangement of psychic materials takes place preconsciously. We should no more blame a person for his/her repression and subsequent scotoma, than we should praise them for spontaneous reflexes. As a reflex action occurs consciously but normally without one’s advertence or choice, so the censorship proceeds without need of our conscious advertence and deliberate control.

The dual role of the censor, in sum, is the pre-conscious schematizing of psychic materials. This schematizing is influenced by conscious fears and desires, including the rational desire to know. The censor provides the epistemological pre-condition both for understanding and for the blind spots that interfere with understanding. The paradox outlined by Sartre of a repressive censor that is both aware and not aware of the same thing in the same regard has been shown to be merely apparent. The solution to the puzzle requires a philosophic standpoint that is non-reductive and a view of consciousness that is non-homogeneous.¹ A non-reductive standpoint views the psychic level as distinct both from its underlying neurological/chemical levels and from its supervening intellectual/rational levels. A non-homogeneous view of the nature of consciousness recognizes that consciousness has different qualities on different levels of conscious intentionality. Perhaps surprisingly, the concept of the censor serves as a differentiating mark of fundamental philosophic standpoints.

Bibliography


The Problem of the Censor


Thomas Aquinas. *In II de Anima*, lect. 13, #398.
Like “democracy,” the word “pragmatism” is “essentially contested” in W.B. Gallie’s phrase, and there is no agreement about what it means or about whether it can be defined at all. When Arthur O. Lovejoy distinguished "thirteen pragmatisms" in 1908, the British pragmatist F.C.S. Schiller welcomed the idea that "there are as many pragmatisms as there are pragmatists." A century later, after the ideas of C.S. Peirce, William James, and John Dewey have been further developed by later 20th century philosophers and recently revived by Richard Rorty, there are many more pragmatists and many more kinds of pragmatism and neopragmatism. Philosophical theories that have been associated with pragmatism—e.g. radical empiricism, naturalism, historicism, pluralism, and anti-foundationalism—are not always consistent with each other, and John Patrick Diggins describes contemporary pragmatism as having a "split personality" because of the internal disputes within the movement. According to Giovanni Papini, "Whoever gives a definition of pragmatism in a few words would be doing the most antipragmatic thing imaginable," because its dominant feature is recognition of the flexibility of theories and beliefs. The intention here is not to attempt to define pragmatism but to identify the least common denominator of its many different varieties. I argue that the most fundamental characteristic of pragmatism is a particular type of temperament that is best described in terms of adjectives and adverbs, rather than nouns ending in “ism.” Following William James’ dichotomy between tough-minded and tender-minded temperaments, I distinguish two different styles of thinking—the flexible and the fixed. The philosophical tradition has tended to emphasize the fixed habit of thought over the flexible (with the notable exception of Socrates). I try to show that adopting the pragmatic temperament would have important social, political, and educational implications as a corrective to the dominant habit of mind.

Most interpreters assume that there must be a common core of doctrines

4Giovanni Papini, Pragmatismo (Milan, 1913; 3rd. ed. Florence, 1927), 75; 91; quoted in Wiener, 552; 563.
that all pragmatists accept, or at least some beliefs that they share in the way of "family resemblances," but this assumption can be challenged on the grounds that pragmatism was never intended to be a "philosophy" in the usual sense of the word. According to James, pragmatism "stands for no particular results" and is "only an attitude of orientation," but its triumph would mean "an enormous change in . . . the 'temperament' of philosophy."\(^1\) Dewey describes pragmatism as a "disposition . . . as obnoxious to ultimate philosophic truth as it is repellant to certain temperaments" which "discourages dogmatism and its child, intolerance."\(^2\) This disposition recognizes that truth, like health and justice, is "adverbial" in the sense of being a modifier of special cases of action. "How to live healthily or justly [and we might add, truly] is a matter which differs with every person. It varies with his past experience, his opportunities, and his temperamental and acquired weaknesses, and abilities."\(^3\) By saying that truth is not a noun but an adverb, Dewey means that it is an ongoing process rather than a fixed and final goal.

More recently, Louis Menand has described the essence of what the pragmatists taught as the belief that "ideas should never become ideologies--either justifying the status quo, or dictating some transcendent imperative for renouncing it."\(^4\) Richard Bernstein argues that pragmatism should not be interpreted as "a set of doctrines or even a method." He writes, "We can best appreciate the vitality and diversity of this tradition when we approach it as an ongoing engaged conversation consisting of distinctive--sometimes competing--voices."\(^5\) Bernstein seems to forget his own admonition, however, when he lists a series of "isms," including antifoundationalism, fallibilism, and pluralism, in identifying the themes that characterize the pragmatic style of thinking.\(^6\) Similarly, Rorty somewhat inconsistently relies on "isms" to characterize his version of neopragmatism. While endorsing Dewey's view that philosophy is a "social hope" rather than a form of knowledge, Rorty describes his revival of pragmatism as an attempt to weave together "Hegelian historicism" with a Wittgensteinian "non-representationalist account of language."\(^7\) The prevalence of the language of "isms" in discussions of pragmatism has contributed to the view that it is a collection of theories rather

---

\(^6\)Bernstein, 326-329.
than a way of thinking that is compatible with any number of different philosophical approaches.

The widespread tendency to think of pragmatism as a "school of thought" that is in competition with traditional philosophies and contemporary analytic methods is an example of a bad philosophical habit that has a long history. Ever since Aristotle criticized Plato for separating the Forms, it has been noted that philosophers have a tendency towards reification, a category-mistake that treats abstractions as if they were substantial entities. Alfred North Whitehead called this the "fallacy of misplaced concreteness," and Jacques Barzun referred to the same error as "misplaced abstraction."\(^1\) One version of this fallacy treats *qualities* of things and *ways* of doing things as if they were actual *things*, which is to confuse adjectives or adverbs with nouns. According to Aristotle, Plato made this type of mistake in thinking that something called "Justice" exists apart from the way in which just individuals and fair systems of government operate. This intellectual move is analogous to inferring from the fact that a dancer moves gracefully that there is a transcendent Form called "Grace" that she miraculously received from the goddess of beauty or that she is graceful by virtue of her knowledge of aesthetic principles. The common interpretation of pragmatism as a philosophical school commits a category-mistake analogous to Plato's in confusing a concrete way of thinking with a collection of abstractions. We can avoid the fallacy of reifying pragmatism by treating it as a habit or style of thought rather than a set of doctrines. To put the point linguistically, it is misleading to try to define pragmatism in nominative terms, as if it were a fixed entity or ideology. The significance of pragmatism is better expressed by using adjectives to modify a habit of mind or adverbs to qualify a way of thinking.

William James remarked that "The history of philosophy is to a great extent that of a certain clash of human temperaments,"\(^2\) and he argued that pragmatism could be the "happy harmonizer" between two types of temperament, which he described as follows:

<table>
<thead>
<tr>
<th>The Tender-Minded</th>
<th>The Tough-Minded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rationalistic (going by 'principles')</td>
<td>Empiricist (going by 'facts')</td>
</tr>
<tr>
<td>Intellectualistic</td>
<td>Sensationalistic</td>
</tr>
<tr>
<td>Idealistic</td>
<td>Materialistic</td>
</tr>
<tr>
<td>Optimistic</td>
<td>Pessimistic</td>
</tr>
<tr>
<td>Religious</td>
<td>Irreligious</td>
</tr>
<tr>
<td>Free-willist</td>
<td>Fatalistic</td>
</tr>
<tr>
<td>Monistic</td>
<td>Pluralistic</td>
</tr>
<tr>
<td>Dogmatical</td>
<td>Sceptical</td>
</tr>
</tbody>
</table>

If we assume for the sake of argument that James is right in thinking that different philosophical approaches are associated with different types of temperament, it should be possible to identify the pragmatic temperament as a

---


An Anthology of Philosophical Studies

style of thinking that is flexible and open to new ideas, in contrast with an ideological way of thinking that is firmly committed to a set of philosophical doctrines. Using James' model, we can articulate the main characteristics of these two types of temperament in the following way:

<table>
<thead>
<tr>
<th>The Flexible-Minded (Pragmatic)</th>
<th>The Firm-Minded (Ideological)</th>
</tr>
</thead>
<tbody>
<tr>
<td>holds beliefs lightly</td>
<td>holds beliefs tightly</td>
</tr>
<tr>
<td>treats beliefs as hypotheses</td>
<td>treats beliefs as fixed principles</td>
</tr>
<tr>
<td>looks to consequences</td>
<td>looks to origins and foundations</td>
</tr>
<tr>
<td>accepts uncertainty</td>
<td>seeks certainty</td>
</tr>
<tr>
<td>recognizes ambiguity</td>
<td>makes sharp either/or distinctions</td>
</tr>
<tr>
<td>welcomes new ideas</td>
<td>stays loyal to tradition</td>
</tr>
<tr>
<td>adapts to change</td>
<td>remains steadfast</td>
</tr>
<tr>
<td>respects different points of view</td>
<td>requires unity and uniformity</td>
</tr>
</tbody>
</table>

These personality types are easily recognized and much studied by psychologists, but, of course, nobody is a pure example of either one. We are all a mixture of both types and switch from one to the other depending on the circumstances. In our spouses, for example, we usually expect steadfast adherence to the vow of fidelity, but we prefer a more flexible attitude when planning holidays with the in-laws. We want our doctors to change their beliefs and practices in the light of new advances in medical knowledge, but we hope that they will remain firmly committed to the Hippocratic Oath. We elect politicians who we think are open to the different points of view of their constituencies, but we wish that they would stand by their campaign promises. Nobody can survive for even an hour without both kinds of temperament, and so, as James points out, "it is a difference rather of emphasis, yet it breeds antipathies of the most pungent character between those who lay the emphasis differently." From the flexible-minded point of view, the firm-minded are thought to be rigid block-heads, antiquated stick-in-the-muds, or authoritarian intellectual Fascists; from the firm-minded point of view, the flexible-minded seem to be rudderless boats, spineless relativists, or dangerous nihilists whose attitude is, as James puts it, "a mere mess of anarchy and confusion... so much sheer trash, philosophically." On my reading, the pragmatists did not deny that both the flexible and the firm habits of mind are necessary to life, but they made the radical suggestion that we should try treating our philosophical theories in a way that is more like the scientist's entertaining of a hypotheses to be tested or the way we plan holidays in a large family than it is like the attitude we take towards our marriage vows or the doctor's commitment to doing no harm.

Although Papini is right that pragmatism cannot be defined in terms of fixed doctrines, three main characteristics of the pragmatic temperament can be specified: willingness to accept doubt and uncertainty, openness to change, and

---

An Adverbial Interpretation of Pragmatism

recognition of a wide plurality of perspectives. In contrast, the firm habit of mind seeks certainty, fixed principles, and a single, all-encompassing point of view. The difference between these two types of temperament cuts across James’ distinction between the tough and tender-minded personalities. Tender-minded people can be firm (e.g. religious fundamentalists and the Pope) or flexible (e.g. Emerson and the "prophetic pragmatist" Cornel West). Tough-minded people can be firm (e.g. dogmatic atheists like Richard Dawkins and Christopher Hitchens) or flexible (e.g. scientists who adopt Karl Popper's criterion of falsifiability).

The most obvious contemporary examples of the flexible and firm habits of mind are President Barack Obama and his predecessor George W. Bush, both of whom would be classified as "tender-minded" in James' scheme, because they are religious believers, but they have distinctively different ways of thinking. Bush is a good example of the fixed type of temperament, because he holds tenaciously to his ideas about good and evil in spite of all obstacles. He believes in "a divine plan that supersedes all human plans" and thinks that "Moral truth is the same in every culture, in every time, and in every place." He describes his religious faith as "a foundation that will not shift," an attitude that helped him as president to "pick a course and not look back." Obama, on the other hand, illustrates the flexible habit of mind in his interpretation of the American Constitution as "a rejection of absolute truth, the infallibility of any idea or ideology or theology or 'ism,' any tyrannical consistency that might lock future generations into a single, unalterable course, or drive both majorities and minorities into the cruelties of the Inquisition, the pogrom, the gulag, or the jihad." He admires the American Founders because they "were suspicious of abstraction and liked asking questions . . . it was their realism, their practicality and flexibility and curiosity that ensured the Union's survival." Obama does not think that his pragmatic attitude requires “abandonment of our highest ideals, or of a commitment to the common good,” arguing that the American constitution’s rejection of absolutes has encouraged the process of analysis and argument that “allows us to make better, if not perfect, choices, not only about the means to our ends but also about the ends themselves.” Obama’s flexibility of mind is further illustrated in his recognition of the limitations of his own pragmatism. He writes that “it has not always been the pragmatist, the voice of reason, or the force of compromise that has created the conditions for liberty.” It has sometimes been “the cranks, the zealots, the prophets, the agitators, and the unreasonable--in other words, the absolutists”--like William Lloyd Garrison, Frederick Douglass, and Harriet Tubman, who “first sounded the clarion call for justice” and “fought for a new order.” He succinctly captures the inevitable irony of the pragmatic attitude by saying, “I am robbed even of the certainty of uncertainty—for sometimes absolute truths may well be

Bush's tenacious adherence to his view of traditional religious principles is admirable and comforting to those who expect their ideas of right and wrong to be absolute and unchanging. On the other hand, Obama's flexible way of thinking is liberating to those who seek freedom from obsolete ideologies in order to develop new solutions to problems. If the crucial issue in debating about pragmatism is deciding between different types of temperament, rather than between systems of knowledge, then the question "Is pragmatism true?" becomes as meaningless and inappropriate as asking "Is gracefulness true?"

The important philosophical question is, "Should we try to think flexibly or should we remain firm in our convictions?" Many have argued that pragmatism entails relativism, but if we interpret pragmatism adverbially, this is a pseudo-problem which does not arise. No rhetorical sleight of hand can conjure up the logical derivation of even a single proposition, let alone an entire philosophical theory, from a way of thinking, any more than a system of aesthetic principles can be deduced from the quality of gracefulness. Adopting the pragmatic temperament does not entail the view that “anything goes” or that all opinions are equally good, because it does not entail any philosophical views at all. A flexible mind can choose between firm commitment and openness to new ideas, but a fixed mentality has only one habitual way of thinking and thus fewer options for adaptation to change.

Empirical research has shown the advantages of teaching openness to uncertainty, change, and different points of view in the training of doctors and pilots. A large NASA study in the 1970’s concluded that “many cockpit mistakes were attributable, at least in part, to the ‘God-like certainty’ of the pilot in command. If other crew-members had been consulted, or if the pilot had considered other alternatives, then some of the bad decisions might have been avoided. As a result, the goal of CRM [Cockpit Resource Management] was to create an environment in which a diversity of viewpoints was freely shared.” As Jonah Lehrer points out in How We Decide, many hospitals now realize that the same decision-making strategies that can prevent pilot error can also prevent medical error and have adopted CRM techniques in their training programs. “The reason CRM is so effective is that it encourages flight crews and surgical teams to think together. It deters certainty and stimulates debate. In this sense, CRM creates the ideal atmosphere for good decision-making, in which a diversity of opinions is openly shared. The evidence is looked at from multiple angles, and new alternatives are considered. Such a process not only prevents mistakes but also leads to startling new insights.”

In How Doctors Think Jerome Groopman argues that most medical schools do not adequately prepare doctors to accept the uncertainty that is inevitable in the diagnosis and treatment of patients. Instead of using exposure to controversy as an exercise in teaching uncertainty, he sees a “culture of conformity and orthodoxy” in

---

1 Obama, 112-116.
2 Jonah Lehrer, How We Decide (Boston: Houghton Mifflin, 2009), 253-255.
3 Lehrer, 255-256.
medical training that does not encourage open-mindedness but educates for "dogmatic certainty, for adopting one school of thought or the other, and for playing the game according to the venerable, but contradictory, rules" that each institution seeks to impose on staff, students, and patients. Groopman concludes that "Paradoxically, taking uncertainty into account can enhance a physician’s therapeutic effectiveness, because it demonstrates his honesty, his willingness to be more engaged with his patients, his commitment to the reality of the situation rather than resorting to evasion, half-truth, and even lies. And it makes it easier for the doctor to change course if the first strategy fails, to keep trying. Uncertainty sometimes is essential for success."

Given that the pragmatic temperament demonstrably improves the performance of doctors and pilots, what might be the effects of adopting it on a larger scale? James describes his hopes for the impact of pragmatism as follows: "See, I say, how pragmatism shifts the emphasis and looks forward into facts themselves. The really vital question for us all is, What is this world going to be? What is life eventually to make of itself? The centre of gravity of philosophy must therefore alter its place . . . It will be an alteration in 'the seat of authority' that reminds one almost of the Protestant Reformation." To many readers James' comparison of pragmatism to the Protestant Reformation has seemed preposterous. Rorty writes, "In an exuberant moment, James compared pragmatism's potential for producing radical cultural change to that of the Protestant Reformation. I would like to persuade my readers that the analogy is not as absurd as it might seem."

The apparent paradox that a philosophical approach with no direct and immediate results could have the potential to bring about an intellectual revolution does not appear so absurd if we interpret pragmatism adverbially as a habit of mind rather than a philosophical theory. There is a chance that teaching the pragmatic temperament might result in a paradigm-shift with important implications both within and outside the academy. Many teachers at every level of education claim to follow Dewey’s idea of “learning by doing,” but few put this idea into practice in their grading policies. Most teachers do not fail students who excel on objective tests but show no evidence of being open to different points of view, nor do they give honor grades to students who do poorly on standardized tests but succeed in developing open-minded and flexible habits of mind. As Dewey recognized, “something hardly less than a revolution in teaching would be worked” if teachers began to see that the true measure of educational growth is not the production of correct answers but the development of the pragmatic temperament.

James, Dewey, and their followers, like Nietzsche, envision a future in which there would be greater individual freedom of thought; unlike Nietzsche, they are democratic in emphasizing the inseparable relationship between strong individuals and a strong sense of solidarity with human communities of

---

2James, “Some Metaphysical Problems Pragmatically Considered,” 86.
3Richard Rorty, Philosophy as Cultural Politics, x.
inquiry. Those of us who advocate self-consciously adopting and teaching the pragmatic temperament do not claim to know that it would be better for humanity in the long run to try out this intellectual and social experiment. We imagine that this way of thinking could lead to a future society in which people are freer and happier, because they will be better able to adapt to uncertainty, change, and a plurality of points of view. I have argued that the most valuable legacy of pragmatism is its clear presentation of the choice between the flexible and the firm habits of thinking as "live, forced, and momentous." If my argument has been successful, teaching the pragmatic temperament will begin to receive serious consideration at every level of education.¹

¹I am grateful to Catharine Forbes, Lucien Frary, Robert Good, Louis Menand, and Guy Stroh for their comments and helpful suggestions.
CHAPTER SEVENTEEN

Merleau-Ponty's Doctrine of Objects

Henry Pietersma

Under this title I want to discuss objectification and what in Merleau-Ponty is intimately associated with it, namely conceptualization. I put my interpretation in the wide framework of the transcendental philosophy that goes back to Immanuel Kant. To be sure, there are important differences between him and Kant, but the general characteristics of transcendental philosophy are clearly found in his writings, specially in his *Phenomenology of Perception*. Thinking in terms of objects, characteristic of common sense as well as science, is for both philosophers a problem of great importance. One may put it in the form of this question: “How does experience of objects come about?” This is the question transcendental philosophy wants to answer. In both Kant and Merleau-Ponty concepts and their relation to temporality play a central role. As the former would have it, a temporal manifold is objectified by the application of transcendental concepts. Merleau-Ponty does not agree that the concepts involved have their ground in pure reason, but are rather modes of temporality. Objects for him are a kind of primordial transformation of time. Let us turn to his doctrine, which is mostly to be found in his *Phenomenology of Perception*.¹ He relates both objectification and conceptualization to temporality, because they involve something like a suspension or interruption of time. They are, in the technical phenomenological sense first introduced by Husserl, a modification of primordial temporality. To get a handle on this difficult concept we have to consider temporality phenomenologically, that is to say, as a fundamental feature of perceptual experience, specifically primordial in distinction from empirical or secondary perception. In the latter kind of perception we view what we encounter in the light of current linguistic meanings and take it as an instantiation of available concepts. Each individual entity is taken simply as another instance exemplifying one and the same concept and for that reason counted as a familiar object. This is what we do in common sense and science.

Primordial perception, however, is quite different and operates with the ideas of appearances and reality. In this latter framework the percipient transcends appearances by taking them precisely as such, namely as appearances of a reality beyond. Merleau-Ponty here uses the idea of active transcendence in the sense of passing over and going beyond. Applied to primordial perception, transcending appearances means that we do not dwell on what is before us but take it precisely as appearances of a reality beyond, which

¹In the parentheses in this paper the first number refers to the original French edition, the second to the English translation.
is not itself in direct view and accordingly called horizon or world. An example is something of which we see only one side but we nonetheless take the thing to have sides we are aware we can see, though we do not see them at the moment. I do not want to go deeply into this matter here, since it is a well-known point in phenomenological doctrine. What is crucial is that we recognize that the movement of transcendence is basic in primordial perception and distinctive of it. Considering it as a search for knowledge, we have to say that there is here no possession or permanent acquisition, but rather a never-ending, on-going, exploration. Even a privileged perception, as when we say that we see something directly and close up – *en chair et en os*, as Merleau-Ponty himself puts it –, is actually more like a moment between past and future, a passing through, rather than a dwelling on something. When we take, for example, a solid and material object before us as having dimensions we do not at present see, we nonetheless take them to belong to the object and are aware that we can move and actually see it.

Because primordial perception has the character of active transcendence, this philosopher emphasizes that it does not offer certainties. Certainty would mean, I take it, that you no longer go further but remain fixed on what you have in front of you. You might say to yourself “This is certainly the way it is; there is no doubt about it”. Primordial perception, however, does not dwell on that but always goes beyond it, into the wider context earlier referred to as horizon or world. As Merleau-Ponty notes, an unchallengeable presence of what we would ordinarily call a sensible object in a privileged perception is still accompanied by an absence. What is perceived is taken as an appearance, not in the sense of a deception or illusion but as a promise that there is more than what you have now. Putting it differently, a thing is always in-a-world and I am able to move in that world. He goes so far as to say that, if something were completely known, it would by that very fact forfeit its status as a real thing.

In a particularly puzzling passage in the chapter on the cogito, he says that *certainty is doubt* (454[369] author’s italics). I am not altogether sure as to why he put it in this very paradoxical way. As he hastens to make clear, however, this claim is not intended to give grounds for philosophical scepticism. As we soon are told, any particular certainty we presume to have acquired is contextual or relative to a framework. This I take to mean that any particular perception takes place within a specific context. Certainty exists only under certain conditions, within a “tradition of thought”, to use one of the author’s own phrases. In other words, what is handed down from the past and taken for granted. This does not create a threat of scepticism, because, thanks to the transcendence operative in primordial perception, error is sure to be discovered, so that we need not fear of remaining imprisoned in falsehood without escape. (50[40]). The movement of transcendence has its beginning in the past and goes into a future for further exploration and possible correction of present mistakes.

In empirical perceptions, however, the forward thrust of transcendence is no longer as strongly felt and what lies before us is taken to be an object with permanence and identity through time and instantiating a certain number of
concepts throughout its duration. If we consider it with the distinction between appearances and reality in mind, we have to say that an appearance, or a group of appearances, has here been conceptualized and construed as an object. This accounts for its permanence and identity through time. Objects are of course familiar to us in common sense and science. According to Merleau-Ponty, however, from a philosophical point of view, they cannot be accepted as real. As he would have it, they are the result of what we just characterized as conceptualization and objectification. Nonetheless, being a transcendental philosopher, it is incumbent upon him to give an explanation of how such a view comes to be held well-nigh universally, which is to say, he owes his readers an account of the modification of primordial perception of the kind I began to give a while ago.

His account of all this is, unfortunately, not given in a consecutive and coherent exposition but rather scattered through his *Phenomenology of Perception*. I cannot discuss all the passages from which I have construed my interpretation – as I did in my *Phenomenological Epistemology* –, but I hope that you will be content with the conclusions I have reached. I already mentioned earlier in this paper Merleau-Ponty’s idea of a suspension of temporality and indicated that this is brought about by conceptualization and objectification, which are therefore, phenomenologically speaking, a modification of primordial temporality, according to the author. He speaks of temporality as summing itself up in a form that does not pass away. This overcoming of temporality (if we may call it that) is accomplished by the introduction of concepts, something already suggested by Kant. The latter held that the temporal manifold is objectified by means of concepts. Concepts, after all, are not intrinsically temporal events or processes. They are the same in every different application, although the individual entities to which we apply them often are not. In a sense concepts are timeless. Kant applied concepts to introduce unity and order into what would otherwise be sheer multiplicity. Their origin, he argued, is mental, inasmuch as they are imposed upon things *ab extra*, not drawn from them by way of abstraction in the traditional manner of Aristotle.

Although it is helpful to recall Kant at this point, we should not overlook Merleau-Ponty’s distinctively non-Kantian perspective. He held that in primordial perception it is not a matter of form being imposed upon sensible content. At this level, he claims, form and content come together. In fact, we might say that on this point he changed his allegiance to Hegel, inasmuch as he holds that concepts are historical and temporal. Hence his talk of temporality as laying hold of segments of itself without letting them slip away into the past. Or again, he speaks of temporality as contracting or compressing itself (*se contracter*). An object, we read, is a “hold on a segment of time” (277[240]), which we already noted is actually itself a matter of temporality. This is the idea that lies behind the difficult notion of a suspension or interruption of time. In somewhat less puzzling formulation, we hold on to the past by means of concepts. That is to say that we interpret what is before us in the living present as an instantiation of a concept, inasmuch as the concept sums up what has
been experienced in the past. This constitutes objects. Concepts of course come in groups or complexes (one concept implying another). Thus they form what we commonly call a conceptual framework, which can also be regarded as a tradition, a framework hailing from the past. The past is in this sense still with us and the concepts involved are for us still valid. What is present and what comes to us out of the future is subsumed under a concept (or a configuration of concepts) and in that way things get interpreted in terms of the past.

As noted, this constitutes for us our familiar world of objects, the kind of world we believe we are surrounded by in what we call common sense and science, particularly the science of nature. Or more precisely, this is, in utmost brevity, Merleau-Ponty’s transcendental philosophical account of that world. Its importance is obvious. For what it discusses or, if you like, describes is nothing other than the world of ordinary experience. We should not overlook, however, the sharply critical edge of what Merleau-Ponty says. It represents a critique that is perhaps expressed most poignantly in the idea of an interruption of the movement of transcendence. According to our philosopher, what has happened in the process we just considered is the withdrawal of things from that movement in which they are really at home. In short, it is an abstraction in the Hegelian sense of that term. To be sure, objectification and conceptualization represent what is actually necessary and unavoidable in dealing with world, but it must not be overlooked and forgotten, as we so often do, according to Merleau-Ponty. For in the terms of his metaphysics (or ontology, if your prefer), objectification strips an entity of its mystery and reality (270[233]). From Merleau-Ponty’s point of view, there is need for transcendental critique as well as a description. With yet another allusion to Hegel, abstraction can only be eliminated and concreteness found, if we are fully alive to the whole from which objectification takes its rise, which for Merleau-Ponty is found in primordial perception, which is not to be identified with our perception of objects.

Objects, then, are not just experienced as ready-made but are the result of objectification, which is a kind of primordial transformation of the context of active transcendence, a halt in its forward movement. It is basically this process of abstraction that is at work in all object perception. We may think, as a matter of common sense and science, that there are objects that in themselves are related to one another in various ways, spatial, temporal, and causal. We may also think that they are causally responsible for our perceptions, but what is taken to be a real, external entity is actually the result of an historical process. The object is the result of a development summed up in the form of a concept. It is, as it were, laid out before us as a conceptual content. Or to put it in another way, objects are the result of objectification or hypostatization of appearances. In empirical or secondary perception (i.e. the perception of objects in their relations), however, the result is abstracted from the process that gave rise to it, but this abstraction is itself forgotten, if I may put it once again in Hegelian terms. As Merleau-Ponty puts it in one place, we forget the important truth that time is the measure of being (381[330]).

So his basic doctrine of objects amounts to the thesis that we have a world
of objects by virtue of a process of objectification. It forms the basis of much everyday experience but also of science in the broad sense. Inquiry or investigation in the broadest sense is impossible without conceptualization. Science clearly cannot get off the ground without conceptually defining its direction and domain. This process of self-definition led to the recognition of so-called truths of reason. The concepts defining a field or framework of inquiry came to be called a priori in distinction from empirical concepts which were called truths of fact. The latter owe their status to what is actually discovered in the course a specific inquiry, but the former are a priori, because they specify in advance the field that is under investigation. In the history of philosophy this has given rise to the very important distinction of two modes of truth as well as the distinction of necessity and contingency.

Other terms used in this connection are “essence” and “existence”. As we know, Husserl, who was a major influence on Merleau-Ponty’s philosophy, spoke of essences with a great deal of emphasis and enthusiasm. And so did Heidegger, although in his published works the latter does not offer much of a detailed discussion of how essences are arrived at. Now Merleau-Ponty follows their lead and approves of such distinctions, but like Heidegger he does not discuss the matter with the detail Husserl did. The explanation of this comparative neglect is probably the fact that for him the most central issues concern the nature and status of primordial perception. The status of a priori concepts and the closely associated idea of essences command less importance.

What is clear is that a priori concepts form a framework or tradition. The best policy is probably to speak here of conceptual frameworks, because concepts play a constitutive role here. An a priori truth depends on the conceptual relation that subsists between the concepts of subject and predicate. It is commonly said that statements of that kind are either necessarily true or necessarily false. As noted before, basic to any inquiry are the concepts that delimit or define the inquiry itself, specifying the domain or scope of the things or phenomena under investigation. Those properties which we ascribe to an object at the start of a specific inquiry and which thus serve as criteria of its identity may also be called essential, because those properties are not derived from the inquiry itself. If we call an objectified concept an essence, statements about such an object are true or false by virtue their conceptual content. Now what our philosopher is primarily concerned to contend is that an a priori truth is specified by the constitutive conceptual framework or tradition of thought. Working with such a conceptual framework or living within such a tradition, there are some things that are true a priori, because we do not question the framework or tradition as such. Essences exist only for those within a specific conceptual framework and are therefore relative to the latter. They are essences only for those working within it. But if for philosophical or other reasons we do not remain within this framework or tradition, we perform something like a transcendental turn. If this is a genuinely philosophical move, as it is for a transcendental philosopher like Merleau-Ponty, one arrives at a fully historical perspective. From his point of view, seeing an essence – a typically Husserlian locution – is not an apprehension of a truly timeless entity but, as he puts it, a
recapture (reprise), an elucidation or explicitation of what has been concretely experienced. One can only speak of it by indicating the framework in which it functions. Consequently, if we step outside the framework marked off by those essences and ask questions about its temporal, historical origins from which it developed, essences and the necessities coming with them vanish and become, as it were, facts and contingences. Certainty becomes doubt. As we noted earlier, according to Merleau-Ponty, perceptual certainties exist, but only if certain questions are not posed and their supporting framework remains in place.

Merleau-Ponty therefore maintains that there is a basic homogeneity between the seeing of essences and the conceptualization we find in empirical and scientific inquiry. Both start with observable matters of fact and seek to interpret them by means of concepts, or, as he puts it, reading a structure into a multiplicity of cases. Fundamental to both sorts of cognition is the consideration of things in the light of concepts, whether as examples of an essence or as cases of an empirical lawlike connection. That in the search for essences the factual status of the phenomena has no justificatory role, while in scientific inquiry it is crucial throughout, is a matter of less importance from the perspective governing Merleau-Ponty’s discussion. (I base this mainly on his lectures published under the title of The Human Sciences and Phenomenology.) The point he considers it crucial to make is that so-called essences owe their being to objectification, which is, in turn, founded on primordial perception. From a philosophical (i.e. transcendental) point of view, then, an essence is as contingent as a fact. There is no fundamental difference between truths of reason and truths of fact.

According to Merleau-Ponty, that essences are thought of as timeless entities is, from a philosophical point of view, a delusion for which he blames language, but I do not want to go into that matter now. A truth represents only the way a certain period of history sums up what is acquired from earlier times. It therefore cannot really be timelessly true. For a statement to be true, it must be indexed, throughout the entire historical movement, to the stage from which it arises, i.e. the time and circumstances of its utterance. Statements about so-called essences, like ordinary perceptual objects, are dependent, because they are born from objectification. Without the inclusion of time, which of course also includes embodied subjectivity, what we commonly call purely objective statements are abstract (in the Hegelian sense). The non-temporal is only the disguise of the acquired. In reality, it is only a case of adhering to the concepts governing an inquiry or, more broadly, those dominating a segment of history that we speak of as a specific tradition of thought. Within that context things have an aspect that is internal and cannot be, as it were, exported to a point of view outside of that tradition without referring to the latter. Only in reference to that context are they necessary, essential, and timeless, rather than contingent, accidental, and for the time being. In short, they only appear to be timeless from a non-philosophical point of view. The

---

1Edited by James M. Edie in The Primacy of Perception and Other Essays.
transcendental is for Merleau-Ponty, as it was for Husserl and Heidegger, a point of view external to the objective orientation of common sense and science. It represents the genuinely philosophical point of view. For Merleau-Ponty that is the point of view from which time presents itself as the measure of being. Timeless beings can only arise by virtue of a halt in the transcendence of temporality.

What are we to say about this doctrine of objects? As I said from the beginning of this paper, it is a transcendental doctrine. That means that it purports to deal with a problem, namely the problem of how we come to have knowledge of entities we think of as being independent of our mind. I think that the view that is here indicated as a problem is basically the view of classical realism going back to Greek philosophy and continuing its history through the Middle Ages into modern philosophy. Not every philosopher in that history subscribed fully to its doctrine, but I think that it remained the generally accepted view until Descartes, who designated it as a problem. As we know, Immanuel Kant associated the problem with David Hume, to whom he presented his transcendental philosophy as a solution. I do not agree with Kant’s answer to Hume. I would be glad to go into the matter at length, but at the moment I will limit myself to a short indication of what I take to be its central position. As I see it, the problem hinges on one’s view of concepts.

Now how do we think of concepts? Or rather, how should we think of them? I would suggest that we think of them as means by which we apprehend the properties and relations of entities external to our beliefs. Our cognitive mind works with concepts in order to acquire knowledge of things. I deliberately formulate my view so as to express what I myself subscribe to. Hume’s problem was that such an apprehension seemed to him to be impossible. And Kant agreed. that, if the objects of our knowledge are truly external entities, we cannot have knowledge of them. Experience of objects, he held, is possible only by virtue of transcendental concepts. Such concepts make objects possible, because they are imposed upon the given by reason. In other words, we have knowledge of objects, as we do in common sense and science, because objects are the result of conceptualization and objectification. Now Merleau-Ponty’s doctrine of objects clearly conforms to this Kantian pattern. To be sure, at what he calls the level of primordial perception he repudiates the Kantian pattern of form and content, but he would have it that this perception is not, strictly speaking, an experience of objects. An experience of the latter kind is a matter of conceptualization. Here concepts are imposed upon what we primordially perceive; alternatively, what we perceive is subsumed under concepts. The doctrine of objects, as I have presented it in this paper, is typically transcendental, which is to say that concepts are constitutive of objects. From the point of view of my realist philosophy, I find Merleau-Ponty’s doctrine unacceptable. In slightly different terms, his doctrine is supposed to be an answer to a problem in philosophical realism which I reject as genuine.
The notion of necessity (and universality for that matter) as it pertains to claims of knowledge in philosophy and science does not command the respect it once did. The criterion of necessity for what counts as knowledge has become largely irrelevant. On the whole it is deemed “metaphysically contaminated” (to invoke a positivist critique of an earlier era); rejecting all concepts not verifiable in nature and thus not useful for science is the way to go. Modern philosophy has since adjusted itself to the demands of modern science, and is more nearly in agreement (at least in many inquiries) with materialistic empiricism, a/k/a physicalism, than with metaphysical systems and its nomenclature. As Jürgen Habermas has recently observed: ‘Modern science has compelled philosophical reason … to break with metaphysical constructions concerning the totality of nature and history.’\(^1\) So, despite some recent dualist and even panpsychist stirrings, materialism still dominates in philosophy of mind. And with rare exceptions, epistemology is done without postulating a priori conditions for knowledge and without allowing any venturing into metaphysics. Thus, metaphysics, with its primary notions necessity, universality, and final causes, belongs largely to a bygone era. As in Kant’s times so in ours, metaphysics has fallen on hard times.

Nevertheless, in this paper I will argue for the indispensability and timelessness of metaphysics as a grounding device for the scientific enterprise, if that enterprise is to fulfill a modicum of *Wissenschaftlichkeit*, an essential systematicity in arriving at demonstrable answers in science. Without a metaphysics that includes the notion of necessity (along with universality and absoluteness), scientific claims lack explanatory force and a grounding in certainty. Without it science builds on shifting sand and deprives human understanding of the critical satisfaction of certainty beyond reasonable doubt.

Kant is the primary theorist in questions of ethics, aesthetics, science and above all metaphysics. His transcendental philosophy is centered on a faculty psychology which carefully articulates the grounding principles of the mind, as

prerequisites for a metaphysics. Even though we have seemingly passed beyond such transcendental ideas, I am encouraged by recent developments which have, ever so cautiously, moved away from strictly mechanistic-computational, and neuro-philosophical explanations of mental functions, toward non-reductive explanations that appeal to some, if implicit, a priori structures of the mind, those not reducible to brain functions. Such a priori features includes our sensuous receptivity, to be sure a severe deterrent for making a case for a metaphysics. Today we seem to have passed beyond such transcendental ideas. To be sure, there are recent developments that adopt one or another aspect of a Kantian approach, but none of them are unqualifiedly Kantian. (More on this later). Patricia Kitcher takes a return to Kant’s philosophy to be a prudent and useful move to combat skepticism and, as she says “it provides an enlightening model of the human epistemic situation.” Similarly, J. McDowell and L. Bonjour appeal to Kant’s notions of “pure” domains, i.e. apriority in the mind’s disparate domains. All of this encourages stepping in this direction and seriously revisit the problem of the necessary connection(s) between disparate domains, in this essay focusing on cause and effect, with the goal of rethinking the legitimacy of metaphysics in the scientific enterprise. I will revisit a famous debate — between Hume’s original challenge to metaphysics and Kant who, awakened by Hume from his ‘dogmatic slumber,’ mounts his famous rebuttal. I hope a rethinking and drawing fresh attention to this debate will be useful for consideration in current debates (which cannot be here fully investigated) to reintroduce the idea of a general metaphysics to advance science.

Let me begin with the “bad news story” of the status of metaphysics in Kant’s time which is not so different in our own. The predicament of metaphysics is pernicious: Human reason keeps pushing further than what is possible to experience about the sense world. Kant describes it well in the opening sentence of the first Preface of the Critique of Pure Reason:

Human reason has this peculiar fate that in one species of its knowledge it is burdened by questions which, as prescribed by the very nature of reason itself, it is not able to ignore, but which, as transcending all its powers, it is also not able to answer (CPuR, Avii).3


Reason is caught in a web of its own metaphysical meanderings, navigating to and fro between the troubled terrains of the sensible and super-sensible worlds, caught as it were between a Scylla and Carybdis of metaphysical perplexities, forever attempting to reach the shore of certainty and final answers. Kant believed that, however fated reason is in its metaphysical wanderings, we cannot be indifferent to it because on metaphysics, he says, rests ‘the true and enduring welfare and happiness of humanity’ (CPuR, A851/B879). Without analyzing this cryptic claim in its entirety, I confine myself to examining a sub-set of humanity’s welfare, the need to establish science on some grounding foundations. One might reasonably ask whether it is even possible to do science without recourse to metaphysics. Depending on how one answers this might contain the “good news” for science and metaphysics.

A prima facie answer to above question is: of course we can do science without recourse to metaphysics! It is not a scientist’s task to ask metaphysical questions. He does not ask WHY something occurs the way it does, i.e. to look for final causes. Why-questions are non-questions in science. An empirical scientist pursues HOW things work, dictated by methodologies of his specialty; he attempts to produce an unbiased record of his findings without waxing philosophical.\footnote{There are exceptions. Scientists such as for example Einstein, Heisenberg, Plank, to name a few, have also been metaphysicians by virtue of asking large universal questions and attempting answers about the workings of physical nature in its very sources. They often attempted to come up with a TOE, a Theory of Everything. The quest for all-encompassing answers to how the world works is the province of metaphysics. The first natural philosophers were metaphysicians by searching for one underlying source that explained “the many,” the multiplicity of the world. Such TOE has never been achieved. Stephen Hawking once famously lamented “Singularity is the scandal for science”. Particularity persistently eludes all universalizing attempts of metaphysics.}

Even if one accepts the proposition, as I do, that science needs recourse to metaphysics, it is important to know what such a metaphysics would look like. For this purpose I will revisit the Hume-Kant debate and attempt to show what useful elements can be gleaned from it even today. As is well known, for Kant to articulate a legitimate metaphysics became his life’s work: to demonstrate the possibility of metaphysics as science for science. Hume argued against such a possibility, not only declaring metaphysics unnecessary, stronger, deleterious to scientific thinking. Kant, equally adamant, declared it vital – essential for any area of knowledge calling itself science.

I begin with a historical summary of the well-known debate before exploring the systematic issue what a metaphysics for science would look like.

BRIEF HISTORY. Hume had thrown a curveball into all metaphysical systems of the 18th C. by his virulent skepticism concerning the certainty of knowledge. This skepticism was not just affecting metaphysical systems of the time, viz. the Wolff-Leibnizian worldview with its scientific and religious optimism, but it effectively undermined the dominant scientific outlook, Newtonian mechanics. Newtonianism was based on one overarching principle
claiming absolute certainty, the Uniformity of Nature Principle, affirming the necessary connection of cause and effect. If Hume’s skepticism prevailed, this principle would be eviscerated in one fell swoop.

Intelectually KANT was nurtured on Continental rationalism, on Wolf-Leibnizian metaphysics, steeped in Cartesianism. Kant was initially trained to become a scientist or mathematician, not a philosopher. But at age 46 he wrote ‘On the Form and Principles of the Sensible and Intelligible World’ which was published as his Inaugural Dissertation of 1770, assuring him a long-sought-after professorship at the University of Königsberg. Hume’s skepticism played effectively into Kant’s own developing disillusionment with the Leibnizian worldview with its elaborate systems of metaphysics, albeit far removed from sense experience. It resembled, so Kant later, an exercise in ‘philodoxy,’ a love of words (CPuR, Bxxxvii). Even before encountering Hume, Kant began doubting Leibnizian thought which held, optimistically, the human mind can know everything, particularly in cooperation with God. For Kant Leibnizian thinking had become dogmatic, claims not justified by experience. It was either Hume’s Treatise of Human Nature or An Inquiry Concerning Human Understanding (the latter shorter work published under an assumed name, since the Treatise had fallen, so Hume, ‘deadborn from the presses’) that came to Kant’s attention and had a thought-changing effect on him as we will now explore. Let's turn to the systematic details of the Hume/Kant discussion.

HUME’S CURVEBALL. Hume’s skepticism about metaphysical concepts enters inconspicuously enough in his “Doctrine of Ideas”, from which follows, with logical precision, his skeptical conclusions. In it Hume explores one of the most famous questions of his time, Where do our ideas come from? Thinkers before him, Descartes, Locke, Berkeley, all had investigated the origin of ideas and concluded, they come either from outside, the world of objects, or they are manufactured from inner sources. Locke, in tandem with his colleague Isaac Newton, had established a classification of ideas in which ideas were ‘atoms of the mind,’ singular identifiable entities that enter the mind and make impressions. Newton had studied atoms of the outer world while Locke studied atoms of the inner world, mental atoms, ideas. Locke’s classification made use of Newtonian nomenclature: there are ‘simple uncompounded ideas’ that tend to evolve into ‘complex compounded ideas’. From these considerations the famous empiricist motto emerged: Nihil est in intellectu quod non prius fuerit in sensu (nothing can be in the intellect which has not first been in the senses). Hume took this rule to a radical conclusion: ‘All legitimate ideas are copy’d from lively, forceful sense impressions; . . . all other ideas are useless and should be committed to the flames’ (see fn. 6). As Locke so Hume denied the existence of innate, a priori, or uncreated ideas. Hume argued further, ideas that cannot be connected to lively sense impressions are illegitimate. The upshot: We have to give up the notion of causal reasoning as having any a priori status. Hume’s challenge was simply: ‘Where is the impression that gives you the idea

---

“cause”? To reason from an event we experience to a presumed cause which we do not, is based on a bad mental habit that never yields a priori certainty. The idea of a necessary connection between cause and effect is bogus. This effectively sealed the fate of the central thesis of Newtonian science: ‘whatever begins to exist must have a cause in existence,’ and robbed the alleged necessary connection between cause and effect the presumption of absoluteness, a scandal for science to be sure.

KANT’S REBUTTAL. Kant began to think of a way to rescue science from Hume’s ‘wrecking ball’ (Thelma Lavine’s coinage1). Echoing Hume, he posed the central question of the Critique of Pure Reason: ‘How is metaphysics as a science possible?’ (CPuR, B22). To answer this question required a new approach. It called for finding the fixed a priori features in human consciousness that would impose legitimacy on scientific concepts, such as cause, substance, force etc., none of which are given in experience but are needed for the language of science. Kant had appreciated Hume’s logic of reasoning. For Hume the question was not whether or not such concepts were right, useful, even indispensable for knowledge of nature, but whether they could be thought by reason a priori, i.e. did they possess an inner truth independent of all experience?2 The fact is, they are indispensable, and we are obliged to use them in science and in every day life. But the question is, is such reasoning based on [absolute] certainty.

Hume was correct: The idea “cause” just like “God” was not open to experience and was therefore outside the realm of scientific thinking.3 For Kant to deny legitimacy to the concept cause could not be the final word. He set out to rescue the necessity of the necessary connection between cause and effect and to restore it to absoluteness. In this, his effort of a deep analysis of human consciousness, he hoped to establish once and for all, something never done before, a general metaphysics for the sciences.

Kant agreed with Hume: Certain ideas are beyond verifiability in the sense world and thus scientifically questionable, but some ideas are required for science and their legitimacy had to be demonstrated.

It is my contention that Kant’s entire critical corpus, all three Critiques (CPuR, CPrR, and CJ) -- not just, as is generally held, the First Critique -- must be taken as an answer to Hume’s skeptical doubts which had infiltrated all major areas of culture--science, morality and art.4 Significantly, even the external structure of the critical corpus indicates Kant’s preoccupation with and respect for Hume. We see a parallel of the Critiques to Hume’s Treatise of

---

1Thelma Z. Lavine, From Socrates to Sartre: The Philosphic Quest (PBS television series first aired in 1983; now also in book form by same title).
3As indicated logical positivists, more unsparingly than Hume, called such unverifiable ideas metaphysical gibberish. It later turned out positivism’s verifiability principle itself was borrowed from metaphysics; thereafter positivism fell into decline.
An Anthology of Philosophical Studies

Human Nature. The Treatise is organized in three books. Book I, ‘Of the Understanding,’ treats space and time, as does Kant’s First Critique in the “Transcendental Aesthetic.” Hume connects Book I and III via Book II of the Treatise, ‘Of the Passions.’ This parallels Kant’s similar organization to connect theoretical and practical reason in the First and Second Critiques, respectively, through the passions, the aesthetic feeling component of the mind analyzed in the Third Critique, Critique of Judgment. As Hume so Kant insists on the legitimacy of feeling in knowledge acquisition. Hume’s Book III of the Treatise, “Of Morals”, parallels Kant’s moral theory in the Second Critique, Critique of Practical Reason.

Kant’s central answer to Hume was to demonstrate the necessary connection between cause and effect. But the end result achieved more: it demonstrated necessary connections between other disparate domains, sensibility and understanding, freedom and necessity, the conditioned and unconditioned.

Let’s detail what is at stake in Kant’s response to Hume as he overcomes the grip of skepticism and achieves an overall comprehensive vision of metaphysics. The combined Prefaces (1781 and 1787 editions) and the Introduction of the CPuR will provide the outline for this paper.

The 1781 Preface laments the fate of reason and metaphysics, the ‘fallen queen of the sciences,’ which was brought down, so Kant, by two culprits: British empiricism and Continental rationalism. Concerning the first, Kant cites the ‘celebrated Locke’ and his ilk, a ‘species of skeptical nomads who despise all settled modes of knowledge and reduce human understanding to physiology’ (CPuR, A9; Kant’s italics). Kant was familiar with Locke’s theory of ideas, i.e. ideas are atoms in the mind, thus physiological in nature. Locke’s theory, we note, is consistent with contemporary physicalism, the mind reduced to brain functions. The second culprit to bring metaphysics in disrepute was Leibnizian dogmatism, a metaphysics of pure thought far removed from all possible experience, practiced mainly as “philodoxy”, love of words (CPuR, Bxxxvii).

These two 18th century dominant Weltanschauungen had contributed to the demise of metaphysics, either by denying it any status (Hume) or inflating it with illusory claims (Wolff-Leibniz). Kant lamented the deep indifference to all settled knowledge that gripped philosophy of his time, in contrast to the sciences which flourished and had made steady progress (CPuR, Axi).

Kant asked, What is to be done? He answered, [There must be] a call to reason to undertake the most difficult task of all, that of self-knowledge, to institute a tribunal to assure to reason its lawful claims, dismiss all groundless pretensions, in accordance with its own eternal and unalterable laws … This tribunal is no other than the critique of pure reason.’ (CPuR, Axi-xii; Kant’s itals.).

‘Critique’ is Kant’s technical term for analyzing the mind for its a priori principles, those independent of the senses. He believed his critical project, once completed, would answer Hume as well as providing a comprehensive
The Necessity of the “Necessary Connection” in Kant and Hume: Reflections on an Old Challenge to Metaphysics

general metaphysics for the sciences.

In the Second Preface we find Kant’s famous claim that his philosophy is a Copernican(-like) revolution, one that must rethink altogether the workings of our mental apparatus and what judgments follow from them.

The Introduction to CPuR features Kant’s answer to Hume. He begins by placing himself within Hume’s conceptual parameters, then launching his counter-argument. As we have seen, in the First Preface Kant had queried why humans are so hell-bent on going beyond possible experience, i.e. doing metaphysics, and why they keep pondering questions for which there are no satisfactory answers, transgressing the bounds of knowledge. For Hume this tendency to go beyond the testimony of the senses was a habit of human nature. Kant agreed with the primary rule of empiricism: ‘There can be no doubt that all our knowledge begins with experience (sense knowledge)’ (CPuR, B1). But immediately we read his famous caveat: ‘But though all our knowledge begins with experience, it does not follow that it all arises out of experience’ (italics supplied). This is not just a coy mincing of words (my initial reaction in an undergraduate course), but Kant insisted there is something prior, more pristine, imbedded in the mind than the first contact with the sense world. We are endowed with certain fixed features of the mind that are lodged in us pre-logically, pre-linguistically, independent of the senses.

To recall, Hume (and Locke) said, sense impressions are the first and only content written on the tabula rasa of the mind. For Kant impressions are merely the occasion providing the stimulus to excite the deeper dormant faculties in us. These pre-reflective powers go to work on the sensations and organize them through the spatial and temporal forms given a priori, making them ready to be subsumed under an entirely different faculty, categorial understanding, to produce conceptual knowledge. Our faculty of understanding, Kant says, subsumes pre-sifted sense data and translates it into conceptual knowledge.

How does Kant establish the necessary connection between cause and effect? Utilizing Hume’s concepts in Section IV of the Inquiry, ‘Skeptical doubts concerning the operations of the understanding,’ Kant renames these for his purposes. In Sec. 4 of the Introduction, titled ‘Distinction between analytic and synthetic judgments’ (CPuR, A7/B11), Kant uses Hume’s distinction between ‘relations of ideas’ and ‘matters of fact’ renaming them ‘analytic judgments’ and ‘synthetic judgments’, respectively. For Hume these two operations of the understanding exhaust all our reasoning. Relations of ideas are those in algebra and when we manipulate concepts in propositions, viz. ‘bachelors are unmarried males’. Such judgments are intuitively certain, universal and necessary. For Hume only these had a priori status. Relations can be between numbers or concepts in linguistic relations, viz. ‘2+2=4 or ‘the whole is the sum of its parts’. Kant renames these as ‘analytical judgments’ because the subject term of a judgment is analyzed out in the predicate term, establishing an identity between subject and predicate, thus achieving necessity and universality, albeit trivially. No new knowledge is gained in analytic
judgments; like relations of ideas they are mere tautologies. Kant’s example of an analytical judgment, ‘all bodies are extended,’ is self-evidently true; we know it a priori; a body that exists in space must necessarily have extension; it cannot be otherwise but it does not expand knowledge.

Hume’s second operation of the understanding, ‘matters of fact’ become Kant’s ‘synthetic judgments’ (CPuR, A7/B11). Synthetic judgments, while expanding knowledge are, as Hume argued correctly, contingent and not logically necessary nor universal. The predicate term lies outside the subject term. Such judgments are explicative, a posteriori, but not absolutely certain. Kant’s example, ‘all bodies are heavy,’ is synthetic, based on subjective experience; ‘heavy’ is not necessarily, logically contained in ‘body’.

Hume’s skepticism is most radical as he drives these distinctions to a logical conclusion. The proposition ‘every effect has a cause’ (‘everything that happens must have a cause’) is synthetic since the concept ‘cause’ lies outside the concept ‘effect’. There is no logical necessity linking cause and effect. Kant reluctantly agrees: ‘cause’ signifies something entirely different from ‘that which happens’ (CPuR A9/B13). However, this logic reduces the Uniformity of Nature principle of Newtonian physics to mere contingency, denying it necessity and absoluteness. No logical necessity can be assigned to causal reasoning; neither Hume’s matters of fact nor Kant’s synthetic judgments have a priori status. At this juncture the scandal of induction is complete. Inductive reasoning as employed in science cannot establish certainty.

Kant attempts to reverse this unacceptable conclusion. He sets out to turn the tables on Hume and to establish a metaphysics that could provide a foundation for science, in particular for the Newtonian worldview. In what follows I will recount in broad strokes Kant’s grounding for Wissenschaftlichkeit – a metaphysics based on the foundational character of the human mind.

Thus far we have seen, Kant used Hume’s concepts of the two Operations of the Understanding, but providing new conceptual terms for his rebuttal. Kant poses a curious question: Is it possible to achieve a priori status for synthetic judgments (Hume’s matters of fact)? In Sec. vi, “The General Problem of Pure Reason” Kant poses the question: ‘How are synthetic judgments possible a priori? (CPuR, B19). Hume had shown that matters of fact are never a priori, nor it seems are synthetic judgments, by the same logic. Kant’s Copernican Revolution in thinking was to show that Hume’s operations of the understanding do not exhaust all possible judgments having scientific worth. There is a third category, viz. the possibility of synthetic judgments a priori, which are certain and independent of experience as utilized in science. (Not all synthetic judgments will have a priori status, only those or in particular those that are demonstrably certain and for which we require universality in science.)

Two questions demand an answer: The first and most crucial, ‘where does Kant locate the a priori conditions that make such judgments possible?’ The second, ‘how in the end does answering the first secure a metaphysics for science?’
(1) To follow Kant through the labyrinth of his architectonic thinking in all three *Critiques* (see my earlier claim) is out of our reach here. An outline must suffice. Kant acknowledges the radicality of his ideas: “I am introducing a Copernican Revolution in my philosophy” as revolutionary as when Copernicus introduced his new theory of astronomy (*CPuR*, note a, Bxxiii). Kant’s project is comprehensive, overarching and requires nothing less than a complete inventory and analysis of the fixed structures of the mind. Human consciousness is comprised of three independent faculties, a fourth is added in *CJ*. Thus, not only did the critical project require a comprehensive analysis of these powers, each with their own structures, but he discovered (late in his critical work) that the power of judgment was, not a “peculiar power” attached to the understanding (as argued in the First Critique), but a necessary reconciling faculty connecting the first two Critiques. (See my earlier claim that a complete answer to Hume had to await completion of the *Critique of Judgment*.)

Kant’s rethinking of the meaning of apriority becomes a factor of utmost importance to reconcile the mind’s unity and overcome critical impasses. The question, how are a priori judgments possible altogether, must be answered in all three areas where *Wissenschaftlichkeit* is claimed, whether in natural science, morality or art. Kant admitted, to work out an answer for the mind’s apriority required a most arduous effort. As indicated his inquiry extended over the whole critical project and the demand, always and everywhere, to find identifiable a priori principles in each power added to the difficulty of the critical project. For the purpose of this paper, I limit my focus on the way in which Kant demonstrated the necessary connection between cause and effect, the possibility of synthetic judgments a priori for science.

Kant’s understanding of a priori is different from thinkers of his time, indeed also from those of our time. Kant defends a consciousness-a priori as opposed to a language-a priori in analytic philosophy. For Hume (and analytic philosophers in general) apriority refers to semantic, linguistic analyticity, to achieve an identity of concepts. By contrast, Kant searches for an apriority different in kind, for the pure, ontologically fixed, pre-logical conditions lodged in our faculties. For our purposes we focus on two: In *CPuR* the faculty

---


2. For Kant the mind’s unity involves self-knowledge, as a “transcendental unity of apperception,” possible only through critique (*CPuR*, A107ff).

3. “How are *synthetic* judgments possible a priori?” was focus of *CPuR* to save science; “How are *moral* judgments possible a priori?” to save morality was focus of *CPrR*; while “How are *aesthetic* judgments possible a priori?” was the guiding question of *CJ* for standards of art and teleology for science.

of sensibility is governed by pure forms of space and time and faculty of understanding is governed by pure categories of cognition.\footnote{1}{The faculty of the will is ruled by a priori freedom in \textit{CPrR}; the faculty of judgment by a priori purposiveness in \textit{CJ}.} How does Kant answer Hume to show the possibility of synthetic judgments a priori? Our faculty of understanding is equipped with pure categories, one of which is ‘cause and effect’ (\textit{CPuR}, A80/B106). Cause and effect reasoning is an a priori, pre-logical category inhereing in us. We think ‘cause’ and link it necessarily to ‘effect.’ But by itself the category of cause and effect cannot perform any tasks. The power of sensibility must first supply the intuitions pre-formed by spatial-temporal conditions. Sensibility intuit the given world but cannot \textit{think} or \textit{understand} it. Kant says famously, ‘Without sensibility no object would be given to us, without understanding no object would be thought. Thoughts without content are empty, intuitions without concepts are blind’ (\textit{CPuR}, B75/A51). For example, the perceptual intuition ‘burning fire’ is the occasion by which the understanding’s ‘organizing spontaneity’ subsumes or connects the intuition under the category ‘cause and effect.’

Kant’s answer to Hume is now clear: The operation by which the mind connects an observed effect to an unobserved cause is not an idle, arbitrary habit or a thoughtless custom, but a necessary connection human consciousness performs a priori because of its pre-existing mental conditions.

\footnote{2}{We cannot treat here the notoriously difficult question of Kant’s “transcendental deductions” that purport to show how the powers of the mind cooperate in precise detail to advance knowledge. The independence of the faculties from each other means, so Kant, “they cannot exchange their functions” (\textit{CPuR}, B75/A51f). Thus, a cooperation between them, ‘deduced transcendentally,’ needs a demonstration. The \textit{Critique of Judgment} must do duty for this critical task; judgment provides the link between two disparate faculties unable to exchange their functions, yet cooperating to yield judgments of knowledge. The validity of Kant’s ‘transcendental deductions’ are, to this day, a hotly debated conundrum.}
CONCLUDING REMARKS. The claim about fixed and unchangeable features in human consciousness, capable at the same time to provide the grounds for a general metaphysics, is not widely shared. To be sure, a theory for the ontologically prior features of the mind, essentially dualistic in nature, is rejected by physicalist theorists from Democritus onward. Now if we look at the contemporary scene we get a series of good news-bad news stories. The good news (from a Kantian perspective) is that thinkers such as Joseph Levine, David Chalmers and Frank Jackson have questioned whether materialism can do justice to the nature of consciousness. The bad news is that their understanding of consciousness is not Kantian: it has nothing to do with a priori conditions for knowledge. Then again, the good news is that thinkers such as Noam Chomsky and Jerry Fodor have postulated genetically hardwired conditions for language and cognition. The bad news is that they insist that this must be understood within the framework of a computationalist-materialist theory of mind. And the good news is that, as a rare exception, Alvin Plantinga has developed a naturalized epistemology that (he insists) entails a non-materialist metaphysics. The bad news is that he does not lay down a priori conditions for any knowledge whatsoever. There is, so far as I can tell, no unqualified Kantian good news story. However, there is, to date, no recent theory of the workings of the mind that has emerged as complete, thorough (and pleasing), as Kant’s demonstrations of how the mind works. Therefore, taking his complete analyses of all the mental powers seriously, in terms of demonstrating the a priori guiding principles in each that yet allow a necessary connection between them, can do much (in the absence of better theories) to serve as a general metaphysics for all sciences.

Bibliography


1Objections to Kant’s metaphysics with respect to “all knowledge”, including modern science, i.e. how Kant’s metaphysics can handle relativity theories in modern physics are well-known. For a very thorough treatment of these questions, see Michael Friedman’s work in Kant and the Exact Science (1992). He grapples with the difficulty of the limitations of applying Kant’s metaphysics to the hard sciences. Friedman contends that “Kant decisively rejects Newtonian absolute space” and certain “procedure(s) for determining any privileged frame of reference.” Friedman explores how Kant deals with motion, space and time (by implication causal reasoning) by drawing on Kant’s total scientific corpus including Metaphysical Foundations of Natural Science (1786) and the late Opus Postumum in which Kant attempts to settle the “transition” problem left unanswered in his critical work. Friedman’s chapter, “Metaphysics and Exact Science in the Evolution of Kant’s Thought” is useful in gaining a broad glimpse of these difficulties.

2I acknowledge valuable insights from Emmett Holman, my husband, who brought me up-to-date on some relevant issues in the philosophy of mind and epistemology.
CHAPTER NINETEEN

Rorty and the Linguistic Turn

Timo Vuorio

What Gustav Bergmann christened ‘the linguistic turn’ was a rather desperate attempt to keep philosophy an armchair discipline.
- Richard Rorty, 1991

For all my doubts about analytic philosophy, I think that the linguistic turn was an instance of genuine philosophical progress.
- Richard Rorty, 1995

If there is one phrase by which to describe the peculiarity of the 20th century philosophy, it is the ‘linguistic turn’. That is not to say that language was not an important topic or theme prior to that century or now, but it is fair to say that it has never such a central and recognized role in our understanding of philosophical problematic as it had then. What makes the phrase such a previous century notion is that ‘there is an increasing widespread sense that the linguistic turn is past’, as Timothy Williamson (2004, 106) has recently put it. Philosophers, especially within the analytic tradition – a school of thought that, ironically, was once born along with the linguistic turn – have started to think that the importance of language has been somehow overestimated, of which we have now, in the beginning of a new century, freed ourselves.

Richard Rorty has a unique and rather ambivalent relation not just to the linguistic turn but to analytic philosophy as well. According to a popular (but over-simplifying) view Rorty pursued his early career within this tradition, but broke ties with it afterwards. As far as the linguistic turn goes, the worldwide use of the very phrase, and the very idea it expresses, derives from the anthology The Linguistic Turn: Essays in Philosophical Method Rorty edited and named in 1967. But as the quotes above show – both from the days when his mature ‘neopragmatist’ philosophy was fully developed – Rorty’s own account of the merit of ‘taking the linguistic turn in philosophy’ is rather ambiguous.

What I do here is offer an overall view of Rorty’s position that covers both the analytic and post-analytic phases in his career. The argument proposed here is to show that the phases are thematically linked, and, in fact, provide a natural metaphilosophical development. I will argue that Rorty’s poetized account of philosophy is the ultimate form of the linguistic turn. But for the people who once were promoting the very cause of the linguistic turn – accomplishing the
ideal of ‘scientific’ philosophy – this result is a *reductio ad absurdum*.

**Philosophical Questions as Questions of Language**

Broadly taken, the thesis of the linguistic turn is that ‘philosophical questions are questions of language’\(^1\). Rorty, as we saw, is careful in pointing out that the phrase is originally Gustav Bergmann’s. But it is not solely the phrase but also its very idea we can trace to Bergmann. According to Bergmann (1964, 111; 1967, 34-5), the ‘fundamental idea of the linguistic turn’ is to notice that the senses of words differ in their ordinary and philosophical uses. For example, the proposition ‘physical objects, such as the table in front of me, do not exist’ has a specific philosophical sense that is absurd or paradoxical to common-sense (that is, to the ordinary use of the words). Even though the views of the linguistic philosophers vary on how to interpret the difference, what is crucial is recognizing that the philosophical propositions depending on *meanings* opens up a new territory of investigation. In *The Linguistic Turn* Rorty formulates the basic idea in the following way:

> [P]hilosophical problems are problems which may be solved (or dissolved) either by reforming language, or by understanding more about the language we presently use. (Rorty 1992, 3)

With the functions of ‘reforming language’ and ‘understanding more our present language’ Rorty captures respectively the ‘ideal-language’ and ‘ordinary-language’ schools, a typical way to divide the analytical movement in the mid-20th century. Both camps distinguished the philosophical uses of words from the ‘ordinary’ (or ‘logical’, ‘empirical’, ‘scientific’) use, and privileged the latter by giving only it a status of meaningful discourse. For the first camp, the idea was to enrich or clarify the ordinary usage of language with

\(^1\)There is no agreement as to where we can actually trace the origin of the linguistic turn – a problem, of course, related to what we mean by this ambiguous expression. Some clarifications. (1) If it is primarily understood by applying a ‘context principle’ (as it is argued by Kenny 1995) then the creator was Gottlob Frege in his *Die Grundlagen der Arithmetik* (1884), although it has been suggested (e.g. Quine 1970) that the use of that principle – a meaning of word needs to be studied in a context of sentence – can also be found already in Jeremy Bentham’s way to analyze concepts, such as duties and rights. (2) If the expression refers to ‘logical analysis’ – in the sense of mathematical logic being a methodological guide of philosophy – then the inventor title goes to Bertrand Russell and his thesis that ‘the essence of philosophy is logic’ (see Russell 1949, 42-69). Rorty (1979, 257-258) also seems to support this view. (3) But if the expression connotes the thesis that ‘all the problems of philosophy are problems of language’ (as we have supposed here, and which I take to be the most popular sense of it), then the proper articulation of this thesis can be found in Ludwig Wittgenstein’s *Tractatus* – as he wrote ‘Alle Philosophie ist “Sprachkritik”’ (Tract. 4.0031; Wittgenstein 2003, 30). See Hacker (1998; 2007) to back up this view. This account is also supported by Bergmann (1967, 30-31) and Rorty (1992, 3). However, according to Hacker, the linguistic turn in this sense was not completed until in Wittgenstein’s later philosophy of *Philosophische Untersuchungen*; closely related Oxfordian philosophy of ‘ordinary language’ can be interpreted either as a second wave or as a final phase of the linguistic turn.
formal device in which the problematic philosophical senses would be settled. For the second camp, the idea was to clear out the ordinary use from the deep-rooted philosophical confusions in using words.

What does interest us here is not the division but the basic setting: there are (a) philosophical problems on one hand, and (b) a new methodology on the other. This might sound trivial nowadays, but surely that was not the case during the first half of last century. Bergmann actually struggled to prove that the problems the philosophers were dealing with, were actually the ‘old ones’, that is, the traditional metaphysical and epistemological issues concerning, say, the issue of realism and idealism. This is not easy to see because, as Bergmann (1967, 32) notes, ‘both questions and answers are so reinterpreted that they have changed almost beyond recognition’. For example, the rival doctrines concerning the foundational check-point for the meaningfulness of scientific language (phenomenalism/physicalism) have with them, unnoticed by their supporters, an implicit ontological commitment to conflicting views about the fundamental nature of reality (idealism/materialism). To have an articulated view on the latter issue, and thereby contributing to the traditional conversation, would be a sign of ‘bad metaphysics’ which many linguistic philosophers tried to get rid of. For many of them, the thesis was more of inventing a new hot topic to study, than, as Bergmann sees it, having a ‘radically new’ way of ‘approaching the old questions’ (ibid, 32). The purpose of many ideal-language philosophers, such as Rudolf Carnap, and many ordinary-language philosophers, such as J.L. Austin, was to have new promising projects (e.g. intensional semantics, a speech act theory). Ironically, it was not until seeing the continuity of the recent with the past when the talk of metaphilosophical progress could arise with a justification. It was, for example, not until the late 1950’s when the similarity in the agendas of Tractatus logico-philosophicus and Kritik der reinen Vernunft was clearly seen ¹.

It is this exclusively Bergmannian sense that leads to Rorty’s view in seeing if any metaphilosophical progress is gained by applying linguistic tools in philosophy. The Linguistic Turn, even recognizing some substantive difficulties on the path, is still full of optimism. But 12 years later, in Philosophy and the Mirror of Nature, whatever optimism there once was, has been replaced by pessimism. The linguistic turn, he says there, was an ‘attempt to produce nonpsychologist empiricism by rephrasing philosophical questions as questions of logic’ (Rorty 1979, 165), but nothing was really gained. By taking language as the primal concern, there seemed to be some short-term success in the reflection of traditional issues – refuting the epistemological skeptic, the realism/idealism debate, the mind/body-problem – but those were based on presuppositions – e.g. the criterion of meaningfulness via verification – that by the course of analytic philosophy were soon to be abandoned.

We can illustrate the change in Rorty’s stance by taking his own account in regard to the mind/body-problem. For the linguistic philosophers – from Carl Hempel to Gilbert Ryle – the issue had been formulated into concerning the

¹The first systematical work to study the transcendental nature of Tractatus, and thereby seeing the Kantian roots in Wittgenstein, was by Erik Stenius (1960).
nature and status of the problematic ‘mental-talk’ – the terms referring to psychological processes and states – in respect to non-problematic physical or behavioral terms. According to so-called identity-theory, the mental processes are brain-processes (e.g. the sensation of pain = the firing of C-fibres) and the identity relation is that of reductive translation. Rorty (1965) turned the issue into that of reporting use of mental terms, and of their referential relation to a certain scientific explanation (e.g. what people call “the sensation of pain” = the firing of C-fibres). Following Quine’s critique of reduction and ‘analytic’ explication, he argued that the mental-talk is actually referring to brain-states, thereby the mental-talk is not reduced but eliminated, which, strictly speaking, amounts to that there are no mental entities such as ‘sensations’. When people talk about ‘pains’ they actually are referring to brain-states. Rorty interpreted this as a victory for the traditional metaphysical doctrine of materialism.

However, during the ‘70s, Rorty’s account changed. He realized that the referential base of his version of identity theory was privileging the descriptive and predictive ‘vocabulary’ of empirical science, a move that was based on nothing else than being caught up in the ‘current philosophical jargon’ (Rorty 1979, 119). Instead of supporting the thesis of materialism it was presupposing it for being a valid argument; so there was no real contribution to philosophically interesting sense of the traditional mind/body problem. The profound effect of this insight into Rorty’s metaphilosophical development has made Neil Gascoigne (2008) – rightly I think – to call it as ‘Rorty’s Kehre’. The result was seen in Philosophy and the Mirror of Nature: the whole idea of searching for a ‘privileged vocabulary’ is a symptom of the tendency of modern philosophy to find an accurate way to represent reality – a project he found optional and suspectible.

This might indicate to us to conclude that, according to Rorty, no progress was gained in philosophy by taking the linguistic turn. But this, as the second quote above hinted, is not the whole story. In a retrospective essay to The Linguistic Turn, written 25 years after its release, Rorty illuminates the case in a following way:

[I]nsofar as the linguistic turn made a distinctive contribution to philosophy I think that it was not a metaphilosophical one at all. Its contribution was, instead, to have helped shift from talk about experience as a medium of representation to talk of language as such a medium – a switch which, as it turned out, made it easier to set aside the representation itself. (Rorty 1992, 373)

We can summarize two different points Rorty positively asserts in this issue:

(a) The linguistic turn is an attempt to transform traditional philosophical problems that once were about experience or consciousness into concerns of language and meaning.
(b) If philosophical questions can be reformulated as or reduced into questions of language, then the concerns and views we have of language (or meaning) reversely affect the views of philosophical questions.

While the first rearticulates the original thesis of the linguistic turn, it is the second one – let us call it the ‘First Big Rortian Insight’ – that is important here. So the outcome is that if we take an antirepresentationalist or a social practice approach to language – by the example of later-Wittgenstein, Sellars, and Davidson – this has its corollaries to our understanding of philosophical problems.

This point can be illustrated by the feature that is said (e.g. Hildebrand 2003) to distinguish the neopragmatism of Rorty from the classical pragmatism of Peirce, James and Dewey. That is the stance toward the technical applicability of the term ‘experience’ and one’s willingness to take the linguistic turn. Rorty formulates the issue this way:

Dewey’s attempt to set aside the problematic of realism and idealism had involved him in obscure and dubious attempt to see ‘experience’ and ‘nature’ as two descriptions of the same events and in the idea that ‘experience become true’. But philosophers like Davidson, who speak of sentences instead of experiences, have an easier time. (Rorty 1992, 373)

The reference to Davidson’s ‘easier’ task – we skip Dewey’s attempts here – amounts to the way to treat ‘truth’ as a primitive notion and holding the view that the attribution of true beliefs cannot be done without the context of language and of its semantic features (i.e. translation). This procedure rules out the possibility to invoke what Davidson (1984, 189) calls the ‘dualism of scheme and content’, a distinction that Rorty takes to be – as Kantian-rooted formulation of representation-problem – the origin of the realism/idealism-debate. Rorty’s early application of Davidson’s philosophy of language was seen in his criticism of the philosophical sense of ‘reference’ that actually marked a rejection of the doctrine of eliminative materialism (and thereby of having a view in idealism/realism-debate): there is no principal difference between ‘talking about’ mental states and ‘referring to’ physical states, since the latter is just another vocabulary to be ‘talked about’.

Talk of Vocabularies

The talk of vocabularies is probably the most explicit feature in Rorty’s way of taking the linguistic turn. In its widespread Rortian use, as a technical term ‘vocabulary’ is almost analogical to the way classical pragmatists speak of ‘ideas’ or ‘experiences’. Typically, it has more invested in it than what can be expected from the rather relaxed way Rorty seems to use it. Let us first say a
few words of its origin in analytic philosophy, and then a few of its particular
applicability to Rorty’s version of pragmatism.

According to Robert Brandom (2000, 156-157), vocabulary is a ‘Rortian
trope’ that has its roots in Quine’s critique of making a distinction between
changes of meaning and changes of beliefs; there is no way to divide linguistic
practice into concerns of language (structure of meanings) and concerns of
theory (structure of beliefs): the norms that dictate what we mean by our words
are not to be distinguished from the practice of applying those norms, that is,
how we use words when expressing our beliefs. Therefore, instead of even
trying to figure out whether changes happen in our beliefs, or in our meanings
or concepts, we do not separate these issues, but rather prefer to talk about
changes in assemblages of words, that is, vocabularies. The idea that meanings
cannot be parted from beliefs can be called ‘holistic thesis’ (Vision 1990, 87).

Applied to philosophy, the holistic thesis means that the problems we call
philosophical are to be viewed as emerging from an adoption of a certain
vocabulary. It also means that adopting a different vocabulary might have the
result that the problems would not occur anymore. (We will come back to this
issue later.)

The vocabulary-talk is backed up by two constitutive intellectual sources of
Rorty’s version of pragmatism: Darwin’s naturalism and Hegel’s historicism.¹

The first is to say that vocabularies are understood in a Darwinian sense: they
are tools in coping with nature; alike to any natural organisms they are
subjected to time and change, which is to say: they are thoroughly – to use a
favorite term of Rorty’s – contingent. Like Darwin’s idea of natural selection,
these are ‘blind processes’: vocabularies do not have a telos outside of
themselves, in the sense of, say, having a representational relation to non-
represented reality. Each vocabulary is to be pragmatically tested by its own
internal functions; i.e. how fruitful it is, whether it solves more problems than it
raises, etc. Therefore, the function of human condition is to offer new – and
hopefully better, survival-wise – vocabularies to stand the test of time.

The Darwinian naturalism is supported by the other important intellectual
idea. According to Hegelian historicism, the history of culture and human
intellect – Vernunft in its rich German Idealist sense – is a continuous series of
conceptual reinterpretations of human practices and interests. It recognizes
change in seeing human history as a process: the essence of change is the
introduction of a novelty, something never existed before – a new vocabulary –
and which, by the time of next conceptual revolution, will be destined to vanish
to the obscurity of the historical re-descriptions of the past, that is to say in
Rortian words, it will be substituted by a new vocabulary. Therefore, for a
Rortian pragmatist, Hegel is proposing the same recognition of contingency as
Darwin does – Hegel in intellectual history in terms of reason and concepts,
Darwin in natural history in terms of organic creatures and species.

Now think of the mental/physical-distinction. We can see them as different
vocabularies. Taking the pragmatic features we can treat both of them as driven

¹Rorty’s most detailed discussion of Darwin and Hegel’s influence is “Dewey between Hegel and Darwin”, in Rorty (1998).
by different, incommensurable interests in achieving goals. The vocabulary of ‘folk psychology’ serves well its function in our daily businesses. If some day in future people will use the physical jargon for self-description, there has then been occurring a conceptual change that has resulted in people using certain terms instead of others, which might to an extent overlap with the older ones. In that case we have not then made a philosophical discovery concerning a right way to represent reality but that of dropping a vocabulary that has not proven to fulfill its function any longer. This scenario is probably unlikely. Rorty (1991a), following Davidson, supports ‘non-reductive physicalism’: even though the world and each of its portions can be described in physical terms, there are such interests serving such functions that presumably are never to be fulfilled by the resources of physical vocabulary.

The Poetic Turn

Supposedly, what has been said so far is relatively familiar, but now I want to link the discussion to a different theme, and this link, as far as I am aware of, is not recognized yet. This is to say that there is a particular Rortian way to develop the course of the linguistic turn to a direction that actually leaves the traditional realms of analytic philosophy – and pragmatism – behind. This is a phenomenon I will call here the ‘poetic turn’. Taking Rorty’s origin in the analytic philosophy, this might sound surprising, but if one follows the development sketched here, taking this step can be seen a rather natural thing to do.

Let us recapitulate the progression so far. The first step was to view ‘philosophical problems as the problems of language’. This is a step of shifting the target from mind or experience to language: understanding language is a key to handle deep intellectual issues. After seeing the significance of language to philosophy, our account of language has its effect on philosophy. So there are two steps taken so far. If one then, as Rorty does, adopts an antirepresentationalist account of language, the outcome will be that the old problematik – circled around the notion of seeing knowledge and truth as an issue of representation – will be dropped. For Rorty, the outcome of Davidson’s famous ‘no language -thesis’ (Davidson 2005, 107) is to say there is no way to ask traditional representational questions about the language (e.g. how the language fits to facts). The issue is now: what is there left to be done for us philosophers?

It is here where the ‘Second Big Rortian Insight’ comes in. If one sees language as the home of intellectual progress, and if one drops the idea that there is a right way to represent reality – to find a privileged vocabulary – it is a short step to realize that the ‘trick’ lies in using words in new, innovative ways. The explicit form of it is, of course, poetry that, by definition, is a form of literary art that deliberatively by using language – marrying the aesthetic value with apparent meaning – creates new ways of expression. So it is no wonder that from the ‘70s on, Rorty’s writings started both topically and
An Anthology of Philosophical Studies

stylistically reflect to a departure from the boundaries of the typical analytic philosopher. The way he argues within a wide range of themes – that still included the core problems of theoretical philosophy – is not that of justifying one’s views with rational arguments but more of new metaphorical recommendations or ‘redescriptions’, backed up with meta-criticism of traditional philosophical points of view (which, in Rortian terms, are suggesting the existence of a privileged vocabulary). David L. Hall (1994), quite accurately, has called him the ‘poet of new pragmatism’.

Here is to be noted, as it always is with Rorty, that we need to be careful with his ways of using notions, and poetry, with its rather notorious philosophical connotations – think of Plato’s fate for poets in his utopia – is especially one. So what does it mean that ‘philosophy is continuous with art’, as Daniel Dennett defines Rorty’s legacy?¹ The idea of it is best seen in Rorty’s way to contrast and compare the ‘scientific’ with the ‘poetic’, especially spelled out in his Contingency, Irony, and Solidarity (1989). However, to make sense of Rorty’s use of the dichotomy, we should keep in mind Rorty’s peculiar way to use dichotomies. Namely, it belongs to a family of related dichotomies Rorty applies in his argumentation: discovering/making, literal/metaphorical, philosophy/literature, knowledge/hope, etc.

In order to spell it out, let us remember that taking the first step in the linguistic turn is, Bergmann says, to notice that the senses of words differ in their ordinary (common-sensical) and philosophical uses. Rorty – after his Kehre – in no way privileges any ‘ideal’ or ‘ordinary’ language, but the way Rorty (1989, 16-20) uses the literal/metaphorical distinction actually reminds us of Bergmann’s division: only literal uses of sentences are what Ian Hacking calls ‘truth-value candidates’. But the difference between literal/metaphorical use is only that of familiarity. Speaking metaphorically is to ‘use familiar words in unfamiliar ways’. As Davidson (1984, 245-264) argues, the metaphorical uses of words ‘have no meaning’, since ‘uttering a sentence without a fixed place in a language game is’, Rorty (1989, 18) remarks, ‘as the positivists rightly have said, to utter something which is not true or false’.

Now think of the science/poetry -dichotomy. There are two levels of reflection to be separated: (1) the literal or common sensical and (2) the metaphorical or philosophical one. In the first level, the science/poetry-dichotomy can be constructed by using the distinction between ‘finding’ and ‘making’. One can common-sensically contrast science with poetry by their ‘outer’ characteristics: it is appropriate to say that science ‘discovers’, while poetry ‘invents’ things. But this, being as trivial as it is, is not a level Rorty qua philosopher is interested in at all; it is merely the metaphorical idea we should adopt from it, and apply it to the second-level reflection, wherein our actual interests are. Namely, in a second level, while describing ‘inner’ characteristics of science, and seeing it consisting of contingent vocabularies, we can reflect even it being as man-made as poetry is. So in this level of reflection, great scientific innovators – such as Aristotle, Galileo, and Freud – are what Rorty

¹In Dennett’s reminiscences of Rorty, June 18, 2007, Slate magazine (http://www.slate.com/id/2168488).
(1989) calls (by Harold Bloom’s term) ‘strong poets’; they are coming up with conceptual tools – vocabularies – that would redescribe ourselves and our environment with novel, fruitful and successful ways. So in this second sense – a philosophical sense – both science and poetry are poetical, thereby fulfilling the function of making; things and their properties are found only within a larger framework of man-made vocabularies. The idea is not to claim that science is poetic in literal sense, but that both are based on basically similar cognitive processes of creating novelties that is referred by the metaphorical use of the term. What makes the poetry – the actual literary genre of Sappho, Yeats and Goethe – a paradigm of ‘being poetic’ is that there the idea of creating novelties has a value of its own; paying attention to the words and sentences uttered is an explicit trait of the whole game. In other words, it is within the realm of poetry where the features that in other cultural areas are present only implicitly, are present explicitly.

Analogically, the same lesson should apply to philosophy as well. Basically the idea of seeing the poetry as a paradigmatic human activity is a natural consequence for a philosopher who thinks that the linguistic turn, after all, was a step worth taken. If we see the problems of philosophy as the problems of language, then, according to the holistic thesis, it is the vocabularies of philosophy that interests us. If we hold like Rorty (1989, 8-9) does that any vocabulary is consistent and coherent in its own terms – any language-game has a sense if we have given it one – then there is no way to try to argue that there is something fundamentally wrong with it, as the traditional ideal-language or ordinary-language philosophers were trying to do. The only way to dismiss – which is more suitable word than ‘solve’ or ‘dissolve’ – the problems arisen by the vocabulary is to create a new vocabulary that does not raise them. So it is the methods that are stressing the function of shifting the point of view or thinking from a different angle – to ‘induce a Gestalt-switch’ as Rorty (1998, 10) says – that belong to philosophical tools of a Rortian pragmatist.

These are such as proposing an alternative story, offering a new metaphor, composing a Geistesgeschichte; all of them have the goal of creating a new vocabulary: to come up with the terms that redescribe the given phenomenon in a new light.

The ‘poetic turn’ can be illustrated in Rorty’s works since taking the Kehre. The most obvious and well-known case is Philosophy and the Mirror of Nature, where he discusses the topic of knowledge by offering it a metaphor of conversation, to replace the old metaphor of mirror, as the guideline in our understanding of what ‘knowing’ is in the first place. The traditional issues concerning mind/body-problem or realism/idealism-debate are constructed in terms of ‘representationalist vocabulary’ that is guided by the mirror-metaphor of knowledge. However, these issues are dismissed by adopting the conversation-metaphor-guided ‘antirepresentationalist vocabulary’. In his later works, the antirepresentationalist vocabulary is applied, for example, by redescribing the objectivity of science in terms of solidarity and reaching a social agreement (see Rorty 1991a, 21-45).
Closing Remarks

Rorty is not the first to see poetry as an exemplary cultural voice – think of, say, 19th century German romanticism – but what makes Rorty a peculiar case is the rather extraordinary route from which he reaches the conclusion. It is ironical that Rorty is stemming from a tradition that took the Kantian demand of the ‘secure path of science’ rather seriously and even had a teacher (Carnap) who is famous for calling metaphysics as ‘bad poetry’. So it is no wonder that for many – who associate philosophy, by definition, with the ideals of truth-seeking, argument and rationalism – Rorty seem to over-stress the poetic element in his rhetorically rich way of arguing, and talking of ‘metaphor’, ‘strong poets’ and ‘imagination’. I do not have space to justify Rorty’s talk here, even though the remarks I have made should indicate that the difference is not as striking as some of his critics – e.g. Susan Haack – have claimed it to be. The philosophical progress does not happen solely through ‘getting the argument right’, but finding new conceptual tools with which to deal with the old issues, and it is the latter point that Rorty extremely emphasizes. And if we look at the recent suspicion of analytic philosophers towards the linguistic turn, it could be the case that also they are realizing that taking the linguistic turn seriously once and for all, will lead to the landscapes that Rorty – without hesitation and without looking back – was travelling on.

University of Tampere

Bibliography


1Kant was already aware of the possibility of philosophy being poetry-like; in Prolegemona he gives the hint of the nature of metaphysics as “eine Kunst heilsamer und dem allgemeinen Menschenverstande anpassender Überredungen treiben” [wholesome oratory suited to the common sense of man]” (P 278; Kant 1993, 29). Naturally, this has nothing to do with Kant’s own strict scientific demands, but as it turned out to be, some of his first-generation followers, notably Schelling and Hegel, did not find the possibility such a dare.
2See my effort in Vuorio (2010) for that.
3I have tried to situate a Rortian position in the recent debate – e.g. Williamson (2004), Hacker (2007) – over the linguistic turn in the analytic philosophy in Vuorio (2009).
Rorty and the Linguistic Turn

Heroes, 3-34, London and New York: Routledge.


PART 2
Metaphysics, Logic and Philosophy of Language
Determinism, as it is commonly understood in the free will debate, is the position that every event or action, including human action, is the inevitable result of preceding events and actions and the laws of nature. Incompatibilists maintain that free will is at odds with determinism and that the two cannot be reconciled—if determinism is true, free will is impossible. Compatibilists, on the other hand, maintain that our ordinary, folk-psychological notions of freedom and moral responsibility are completely consistent with the acceptance of determinism. In this paper, I challenge the claim that compatibilism reflects our pretheoretical beliefs and I present existing results and experimental evidence in social psychology to argue against the compatibilist thesis. In section I, I briefly spell out the compatibilist position. In section II, I take a closer look at the folk psychology of free will and argue that, contra the compatibilist, recent empirical research by Shaun Nichols, Joshua Knobe and others, reveals that our folk-psychological intuitions are essentially incompatibilist and libertarian in nature. I conclude in section III by examining some additional folk-metaphysical commitments and argue that these further undermine the compatibilist thesis.¹

¹ This paper is an abridged version of the one presented at the 5th International Philosophy Conference in Athens. The focus has been primarily limited to an empirical investigation of the folk psychology of free will. For a more detailed discussion of the compatibilist position and the standard arguments against it—including an expansive discussion of the consequence argument and the various compatibilist replies to it—please see my “Compatibilism and the Folk Psychology of Free Will: Old and New Problems.” That paper also includes a section on the phenomenology of free will, which I argue further undermines the compatibilist position.
number of well-known philosophers have held versions of this position. Classical compatibilism argues that the traditional free will debate is a ‘pseudo-problem,’ the product of a series of conceptual or terminological confusions. It maintains that, when the relevant terms are rightly understood, there is no inconsistency in holding to both free will and determinism. Compatibilists argue that free action is to be distinguished from unfree action not by the absence of causes, as incompatibilists insist, but rather by the type of causes at work. That is, free actions are caused by our desires, wants, or willings, whereas unfree actions are not. As long as the action is caused by the inner psychological states of the agent, and is not externally constrained or impeded, the action is free on this account.

Many critics have argued that this definition of freedom is incoherent and that instead of solving the problem it only camouflages it. The real question, they insist, is whether or not our wants and desires are themselves caused. Since compatibilists have to concede that they are, because they accept or assume determinism, critics charge that it does not preserve real freedom. It is incoherent, they argue, to claim that an action is both free and causally determined. If the inner psychological states that determine our choices and actions are themselves causally determined, how can free will be preserved? Critics maintain that the real issue, so far as the will is concerned, is not whether we can do what we choose to do, but whether we can choose our own choice, whether the choice itself issues in accord with law from some antecedent. For reasons such as these, William James famously labeled compatibilism a ‘quagmire of evasion,’ and Anscombe said it is nothing more than ‘so much gobbledygook.’

Compatibilists, nonetheless, insist that whether or not our inner psychological states are causally determined is irrelevant. What is of real importance, they argue, is that the agent does as he/she wants. The only freedom that matters, according to compatibilism, is the ability to guide our conduct by means of our own wants and desires. Compatibilism believes that this is the appropriate definition of freedom since it accords with our everyday common usage. They maintain that our folk-psychological or commonsense notions of freedom and moral responsibility are completely consistent with the acceptance of determinism. They insists that, in ordinary language, freedom essentially means doing as we want to do, not some libertarian notion of freedom. They also insist that our ordinary conception of responsibility is perfectly consistent with determinism. For compatibilists, when we say ‘Agent A is responsible for action X’ we mean ‘X happened because A wanted X to happen’ (Levin, 2004: 426). What matters, compatibilists contend, is whether or not the agent acted voluntarily, not whether the action is determined by upbringing, heredity, personal psychology, brain chemistry, or the environment. This, they maintain, fits with our ordinary usage, and for most classical compatibilists, common usage is the appropriate criterion for deciding how these notions should be defined (e.g., W.T. Stace, 1952; Levin, 2004; Baumeister, 2008).
The obvious question, then, is whether the compatibilist is correct in assuming that our ordinary understanding of free will is a compatibilist one. Do people generally believe free will is compatible with their own inner psychological states being causally determined? I believe the answer is no! Incompatibilists (like myself) insist that our ordinary concepts of free will and responsibility require the ability to do otherwise. Furthermore, we maintain that our folk psychology is essentially incompatibilist in that it views determinism as being at odds with the existence of alternative possibilities and the agent’s ability to do otherwise. I maintain that instead of being a reflection of our pretheoretical beliefs, compatibilism is a complex philosophical position—one that is defended by sophisticated philosophical moves. As Thomas Pink writes: ‘Compatibilism is not something naturally believed, but something that has to be taught—by professional philosophers, in philosophy books, and through philosophy courses’ (2004: 43). I agree with Robert Kane that:

[Most ordinary persons start out as natural incompatibilists. They believe there is some kind of conflict between freedom and determinism; and the idea that freedom and responsibility might be compatible with determinism looks to them at first like a ‘quagmire of evasion’ (William James) or ‘a wretched subterfuge’ (Immanuel Kant). Ordinary persons have to be talked out of this natural incompatibilism by the clever arguments of philosophers (1999: 217).

Although I believe ‘we come to the table, nearly all of us, as pretheoretical incompatibilists’ (Ekstrom, 2002: 310), I am a philosopher—and philosophers are not always the best guide to our pretheoretical beliefs! In fact, both sides claim to reflect our ordinary, folk-psychological intuitions. How, then, can we move beyond the traditional stalemate? I propose that we examine our folk-psychological intuitions and judgments empirically. The debate between compatibilists and incompatibilists ultimately comes down to who is right about our folk-psychological intuitions; and this, I maintain, is an empirical issue. The recent experimental philosophy movement proposes that we can shed new light on traditional philosophical problems by empirically studying our pretheoretical intuitions and judgments. This approach, I believe, can be especially helpful in the free will debate where both sides lay claim to commonsense. In the following section, then, I will present existing results and experimental evidence in social psychology that reveals that our folk-psychological intuitions are essentially incompatibilist and libertarian in nature.1

1My thesis is that ordinary folk are pretheoretical incompatibilists who believe in indeterminist free will. If correct, compatibilists will no longer be justified in affirming that their theory elucidates our ordinary conception of moral freedom (the freedom required for moral responsibility), or that their definition of freedom corresponds “to what laypersons generally mean when they distinguish free from unfree action” (Baumeister, 2008: 14). My aim here, then, is simply to establish this thesis—not to explore the implications of conceiving of free will in this way. For more on this latter point, see my Free Will and Consciousness: A
An Anthology of Philosophical Studies

The Folk Psychology of Free Will

Philosophers have long argued that our folk psychology is inherently incompatibilist, but only recently has the issue been explored empirically. Shaun Nichols (2004, 2006a), for example, has argued that recent work in developmental psychology on the understanding of agency in infants and toddlers indicates that young children deploy a notion of libertarian agent-causation. Nichols uses existing results as well as his own experimental data to support this claim. Appealing to the work of Meltzoff (1995) and Johnson (2000; Johnson et al., 1998, 2001), Nichols first argues that developmental research indicates that, from a very early age, children regard agents as distinctive and causal—‘They quantify over agents, and they locate in the agent a causal power to produce an action’ (2004: 482). This, however, is not yet enough to prove children are employing a libertarian notion of agent-causation. Children might regard agents as causes and also expect such causal activity to fit comfortably into a deterministic framework. Unfortunately for the compatibilist this does not seem to be the case. Nichols (2004), through some simply designed experiments, reveals that children also regard agents as having the robust capacity to do otherwise. According to agent-causal theories, an agent is an agent-cause of a given action only if the agent could have not caused it, that is, only if the agent could have done otherwise. Nichols’s experiments indicate that children deploy just such a notion of agent-causation.

In one experiment, Nichols placed children in one of two conditions to investigate whether they regard agents as having the capacity to do otherwise.\(^1\) Those in the agent condition (condition 1) witnessed an agent exhibit motor behavior; those in the object condition (condition 2) witnessed an object move. For instance, children in the agent condition were shown a closed box with a sliding lid. The experimenter said, ‘See, the lid is closed and nothing can get in. I’m going to open the lid.’ The experimenter then slid the lid open and touched the bottom. Children in the object condition were shown the same closed box with a ball resting on the lid; the experimenter then slid the lid open and the ball fell to the bottom. The children were asked whether the agent/object had to behave as it did after the lid was open, or whether it could have done something else instead. The results were very clear. Every single child said the person could have done something else and nearly every child rejected this option for the thing. In a second experiment, both children and adults were presented with cases of moral choice events (e.g., a girl steals a candy bar) and a physical event (e.g., a pot of water comes to a boil), and they were asked whether, if everything in the world was the same right up until the event

---

\(^1\)The mean age of participants was 4 years, 10.5 months. All of the children were at least 3 years old. This is important since some of the experiments required participants to answer counterfactual questions—and studies have shown that three-year-olds are good at answering simple counterfactuals questions, and four-year-olds can answer even somewhat complicated counterfactual questions (Harris et al., 1996).
occurred, the event had to occur. In this setting, both adults and children were more likely to say that the physical events had to occur than that the moral choice events had to occur. According to Nichols, this provides evidence that the folk have an indeterminist concept of free choice on which agents could have done otherwise.

A compatibilist, of course, might object that this data is open to alternative compatibilist interpretations. Nichols, however, considers several alternative interpretations of the data and concludes:

[T]he evidence certainly fits the indeterminist interpretation. And there’s not (yet) any evidence against it. On the contrary, the available evidence provides support for the claim that children embrace both claims of the agent-causal account. Apparently children think that an agent is a causal factor in the production of an action. They also seem to think that when an agent produces an action, he could have done otherwise, and moreover, that when an agent makes a moral choice, he was not determined to choose as he did (2004: 488-489).

I agree with Nichols that the data is best interpreted in an incompatibilist way, and to see why let us examine one leading alternative interpretation. According to the standard compatibilist approach, when people claim that an agent could have done otherwise, what they really mean is a tacitly conditionalized ‘would’ statement—i.e., the agent would have done otherwise if the conditions had been different.\(^1\) With regards to the present study, the compatibilist might say that when the children claimed that the agent could have done otherwise, they were only claiming that the experimenter would have done otherwise under different conditions. It is, no doubt, difficult to rule this possibility out completely. Nonetheless, I believe there are several reasons for not favoring this interpretation of the data. For one, the contrast with the thing-case goes some distance to rebutting this interpretation. For nearly all the children said that the ball had to do as it did, but of course, children understand that the ball wouldn’t have touched the bottom under different conditions, e.g. if the experimenter had turned the box over as he opened the lid. If the children were deploying a conditionalized understanding of ‘could have done otherwise,’ one would expect that they would apply it to both agent and thing cases alike. Since they don’t apply it in the thing case, I believe there is at least prima facie reason to think that they’re not applying it in the agent case either.

---

\(^1\)Classical compatibilists maintain that when we say that an agent can (i.e., has the power or ability to) do something, we mean that the agent would do it if the agent wanted (desired or chose) to do it. They believe this type of conditional analysis fits common usage and allows us to see how the power and freedom to do otherwise can be reconciled with determinism. If the power to do otherwise means only that you would have done otherwise if you wanted or desired, so the argument goes, it would be consistent with determinism since it would not require changing the past or violating laws of nature. To say ‘you could have done otherwise’ would only amount to the counterfactual claim that you would have done otherwise, if (contrary to fact) the past (or the laws of nature) had been different in some way.
Secondly, there are serious philosophical problems with such conditional analyses and there is growing agreement among philosophers that this compatibilist move fails to reconcile free will and determinism (see Broad, 1952; Ginet, 1980; J.L. Austin, 1961, 1966; Chisholm, 1964, 1966; Lehrer, 1964, 1968). In fact, some compatibilists have even abandoned conditional analyses, insisting that if compatibilism is to succeed it must find an alternative strategy (Audi, 1974; Berofsky, 1987, 2002). One of the main criticisms of this analysis is that it appears to lead to a regress problem (see Chisholm, 1964, 1966; Lehrer, 1964).

Thirdly, and perhaps most importantly, one could turn the tables and ask: What psychological evidence do compatibilists have that people actually interpret could have done otherwise in this conditional way? The compatibilist provides us with a philosophically sophisticated interpretation of could have done otherwise, but what reason do we have for believing that people actually employ such a notion? It is simply not enough to present a possible interpretation of what might be going on. I contend that the compatibilist has to make a plausible case that people actually view matters in this way. There is, however, little empirical support for this conclusion. Quite to the contrary, Nichols (2006b) has recently conducted a pilot study that suggests just the opposite. In the study, 75 undergraduates were given the following vignette:

On 4/13/2005, Bill filled out his tax form. At precisely 10:30 AM, he decided to lie about his income. But of course he didn’t have to make this decision. Bill could have decided to be honest.

The subjects were then asked to judge whether a sentence sounded right or wrong (on a scale of 3 to -3). One group got the following sentence, modeled on conditional analyses:

Bill could have decided to be honest at 10:30, 4/13/2005, but only if some things had been different before the moment of his decision.

The other group got the following sentence, modeled on unconditional analyses:

Bill could have decided to be honest at 10:30, 4/13/2005, even if nothing had been different before the moment of his decision.

The results were telling. Subjects were more likely to judge that the unconditional sentence sounded right. They gave higher ratings for the sentence modeled on unconditional analyses than for the one modeled on conditional analyses. Although I think further studies here would be helpful, I agree with Nichols that ‘this helps to shore up the philosophical consensus that conditional analyses do not reflect the folk notion of could have done otherwise’ (Nichols, 2006b: 306).
Further support for the claim that people regard free will and choice as indeterminist can be found in experiments Nichols and Joshua Knobe (2007) recently conducted. Nichols and Knobe presented adult subjects with a questionnaire that described both a determinist universe (A) and an indeterminist universe (B), described as follows:

The key difference...is that in Universe A every decision is completely caused by what happened before the decision—given the past, each decision has to happen the way that it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision does not have to happen the way that it does.

After the description was read, subjects were asked, ‘Which of these universes do you think is most like ours?’ The vast majority of subjects (over 90%) answered that the indeterminist universe (Universe B) was most like ours. Nichols and Knobe conclude that, since the only feature of the universe that is indeterminist is choice, the data indicates that people are committed precisely to the idea that choice is indeterminist. In a second study designed to test the folk-psychological notion of moral responsibility, subjects were read the same descriptions of Universes (A) and (B) and were asked: ‘In Universe A [the deterministic universe], is it possible for a person to be fully morally responsible for their actions?’ Most subjects (86%) said ‘no.’ These results are especially telling because they further reveal an incompatibilist theory of moral responsibility.

Nichols and Knobe did, however, find exceptions to this conclusion. They found that there are certain experimental contexts in which subjects do express compatibilist judgments. But with regard to these cases, they write:

Our hypothesis is that people have an incompatibilist theory of moral responsibility that is elicited in some contexts but that they also have psychological mechanisms that can lead them to arrive at compatibilist judgments in other contexts...[The] data show that people’s responses to questions about moral responsibility can vary dramatically depending on the way in which the question is formulated. When asked questions that call for a more abstract, theoretical sort of cognition, people give overwhelmingly incompatibilist answers. But when asked questions that trigger emotions, their answers become far more compatibilist (2007: 105-106).

Nichols and Knobe found support for the hypothesis that, when people are confronted with a story about an agent who performs a morally bad behavior, this can trigger an immediate emotional response, and this emotional response can play a crucial role in their intuitions about whether the agent was morally responsible. On their view:
[M]ost people...really do hold incompatibilist theories of moral responsibility...It’s just that, in addition to these theories of moral responsibility, people also have immediate affective reactions to stories about immoral behavior (2007: 109).

When subjects are offered questions that call for more abstract, theoretical cognition—questions that rely less on emotional reactions to immoral behavior—subjects markedly exhibit incompatibilist theories of moral responsibility. When, for example, Nichols and Knobe asked the abstract question whether anyone could be ‘fully morally responsible’ in a deterministic universe, subjects tended to say ‘no.’ This reveals that the compatibilist claim—that according to folk psychology, moral responsibility is completely compatible with determinism—is simply mistaken. Incompatibilist intuitions about moral responsibility have also been documented and discussed by a number of other philosophers and psychologists.

In a recent study conducted by Monterosso et. al. (2005), participants were asked to complete questionnaires designed to examine the tendency of scientific explanations of undesirable behaviors to mitigate perceived culpability. Participants were presented with four vignettes depicting personally or socially undesirable behavior and were given two different explanation-types for each behavior: physiological and experiential. Monterosso and his colleagues found that physiological explanations of behavior tend to undermine responsibility attributions significantly more than experiential explanations that appeal to the agent’s history or experience. For example, when a physiological (e.g., neuro-chemical) explanation of a particular action was given, subjects were less likely to blame the agent than when experiential explanations were given. In addition, physiologically explained behavior was more likely to be characterized as ‘automatic,’ and willpower and character were less likely to be cited as relevant to the behavior. I predict that if this trend holds, what we should expect to find is that when underlying physiological features are emphasized (e.g., Gregg’s brain chemistry) this will lead to exculpatory judgments more frequently than cases that ignore these aspects and instead focus on what deliberation was like for the agent. The findings of Monterosso and his colleagues definitely point in this direction.\(^1\) Yet how could a compatibilist explain this? In theory, if compatibilism were correct about our folk psychology we shouldn’t find this. Compatibilism maintains that our folk-psychological notions of free will and moral responsibility are compatible with determination by antecedent conditions—which, in theory, should include determination by neuro-chemical brain activity. If free will and antecedent determination are compatible, then why would explanations of behavior based on physiological causes diminish or undermine ascriptions of moral responsibility? The simplest and most obvious answer is that the folk view certain types of deterministic explanations as

\(^1\)Monterosso et. al. write: ‘the stage is set so that advances in the physiological behavioral sciences progressively shrink what is left to attribute to the intentional agent’ (2005: 155).
Compatibilism and the Folk Psychology of Free Will

incompatible with moral responsibility—more specifically, physiological explanations disincline the folk to ascribe responsibility.

If all this research is correct, this is bad news for compatibilism since the whole compatibilist project is an attempt to reconcile causal determinism with our commonsense understanding of free will. I think Vargas says it best when he writes:

First, these results yield something of a victory for incompatibilists. Incompatibilists need not—indeed they generally do not—deny that there are conditions under which we operate with compatibilist notions of abilities. What incompatibilists maintain is free will and moral responsibility (in some important sense(s)) are incompatible with determinism. For compatibilism to be a meaningful position, it must hold that there is no important sense in which free will and moral responsibility are incompatible with determinism. However, if Nichols’ data are correct, contra to what traditional compatibilists have claimed, we really do imagine ourselves to be agents with genuine, metaphysically robust alternative possibilities available to us. Moreover, we really do—at least in moments of cool, abstract consideration—tend to favor an alternative possibilities requirement on moral responsibility. On the plausible assumption that these senses are important—and minimally, they are important for our self-conception—it turns out that compatibilism is dead wrong about there being no important sense in which free will and moral responsibility are incompatible with determinism (2006: 243-244).

If it turns out that our common sense intuitions regarding free will are wholly or largely libertarian and indeterminist, the compatibilist project comes crashing down. No longer can the compatibilist maintain that their (re)definition of free will reflects (or is equivalent to) our ordinary usage, nor can they appeal to our folk-psychological intuitions as ‘proof’ of their position.

Additional Folk-Metaphysical Commitments

In addition to the above results, a closer examination of our folk-metaphysics (our collective commonsense intuitions regarding metaphysical matters) reveals other libertarian commitments. Paul Bloom, a developmental psychologist, has persuasively argued that young children have a number of philosophically ill-mannered commitments. In his fascinating book, Descartes’ Baby (2004), he presents several studies that reveal that a major component of our folk psychology is our belief in dualism—the belief that the mind is fundamentally different from the brain. According to Bloom, the belief in dualism comes naturally to children and this dualistic worldview persists in adulthood. This widespread belief in dualism provides further proof that our folk psychology is inherently libertarian (especially agent-causal) and not compatibilist. Belief in dualism fits nicely with belief in agent-causal powers,
and many agent-causal theories adopt dualist metaphysics (e.g., Foster, 1991; Eccles, 1994; Swinburne, 1986). In many ways, these two beliefs go hand-in-hand. Bloom, in fact, suggests that the belief in indeterminist free will is another piece of this central dualistic worldview.

If compatibilists are going to insist that our folk-psychological intuitions are inherently compatibilist, they should focus not only on intuitions about moral responsibility—which, again, are often incompatibilist—but also take account of intuitions like these. Our folk psychology reveals a robust folk metaphysics, and it would be disingenuous of compatibilists to focus only on those intuitions that support (or sometimes support) their position while ignoring those that don’t. There is reason to think that folk-psychological dualism plays a major role in intuitions and ascriptions of moral responsibility. In the study conducted by Monterosso and his colleagues, we saw that physiological explanations of behavior tend to undermine responsibility attributions significantly more than experiential explanations. Monterosso et al. speculate that this might have to do with an underlying commitment to dualism in ordinary person perception. They write:

\textit{Put together, these three propositions—(1) mind (soul) and body are two separate entities capable of independently affecting human behavior; (2) behavior is voluntary and “owned” by the self only insofar as it flows from the mind and soul; (3) alternative, independently plausible, and non-contradictory accounts of human action can undercut each other—may elucidate why, in our vignettes, when a physiological explanation was given, there was a tendency for participants to view the body as the cause of the behavior, and motivations as less relevant, with the result that the behavior was perceived as less voluntary...More generally, in viewing the mind and body as two mutually exclusive attributional suspects (as opposed to alternative levels of analysis), the stage is set such that advances in the physiological behavioral sciences progressively shrink what is left to attribute to the intentional agent (Monterosso et. al. 2005: 155).}

Given the plausibility of this hypothesis, compatibilists cannot ignore the folk-psychological acceptance of dualism. If Bloom and others are correct, dualism is a major part of our folk psychology, and any account of our folk-metaphysical conception of free will needs to take into account (and make sense of) these other metaphysical commitments. Here compatibilism fails.

\textbf{Conclusion}

The compatibilist/incompatibilist debate has traditionally been a war of intuitions. To help advance the debate beyond its current stalemate, I proposed that we examine our pretheoretical intuitions empirically. I then presented
existing results and experimental evidence in social psychology to argue against the compatibilist thesis. I argued that the findings of Nichols, Knobe, and Monterosso et. al. reveal that our folk-psychological intuitions are essentially incompatibilist and libertarian in nature. Although we might occasionally express compatibilist judgments—especially in cases of moral choice events that trigger emotional reactions—the findings suggest that our more abstract, theoretical reasoning is overwhelmingly incompatibilist. If correct, this yields something of a victory for incompatibilists. No longer can the compatibilist maintain that their definition of free will reflects (or is equivalent to) our ordinary conception, nor can they appeal to our folk-psychological intuitions as ‘proof’ of their position. The only avenue left for the compatibilist is to defend a revisionist account of free will—one that proposes we accept a definition that deviates from our ordinary conception. To do so, however, would be to undermine the very goal of compatibilism. To give up our ordinary conception of freedom (for a revisionist one) would be to admit that what we ordinarily think of as free will cannot be reconciled with determinism!

**Works Cited**


Harris, P., T. German, and P. Mills (1996). ‘Children’s Use of Counterfactual Thinking
Clara’s life up to now fits into a coherent story pattern. Indeed, her story is one that she tells herself. From a very early age she wanted to be a concert pianist and her artistic parents encouraged her in that direction. Born in Sao Paulo in the 1960s, she took private lessons as a child and then went to the local music academy where she won numerous piano prizes. She continued her piano studies in Germany and then in the U.S., where she received a Ph.D. She is now teaching at a U.S. university and concertizing internationally. Her solo and accompanying repertoire is vast, although she has a special love for contemporary classical Brazilian piano music. Clara does have many other interests than her musical career, but with some exceptions these are kept subordinate. When she thinks of various episodes in her life, including both pleasant as well as frightening ones, in most cases she thinks of them primarily in relation to her career, as furthering it or sometimes setting it back. Her romantic involvements, for example, are evaluated by her in this way, as are her traveling experiences. Asked about a specific event in her childhood, she might say, ‘Oh yes, I remember hearing that Villa-Lobos piece for the first time and that early experience still has some influence on my playing today.’

Harry’s overall life up to now has been pretty incoherent. He came from a wealthy New England family, went to a private secondary school in Boston, somehow spent his college years at a Florida university in the 1970s ---where he experimented with sex, alcohol, and some drugs --- and had no idea of what he wanted to do after college. From Florida he drifted into a job overseas, teaching English to North-African students in Morocco. After North Africa he came back to the U.S., lived with his parents and had a nervous breakdown. In the 1990s he had a series of unrelated jobs, but starting in the late 90s worked as a bank teller at Wells Fargo for almost a decade. He lost his Wells Fargo job, however, in the recent economic crisis and moved to Michigan, where he met and married a wealthy widow. Last summer, and now financially secure, he became obsessively interested in recreational biking, often biking over 30 miles a day. Today, Harry can remember much of his past life, although some of the period of his nervous breakdown seems rather blurry. He cannot, however, fit many of these past experiences into a coherent whole. He sees little relation, for example, between his experiences in North Africa and his present biking interests.

Clara’s whole life so far fits into a unified story; Harry’s does not. Clara can tell us her story; and even if she were not a good story teller, we could tell
An Anthology of Philosophical Studies

her story. Harry cannot tell us much of a story about his life and nor can we. In fact when his new wife recently hired a professional to write a biography of Harry (God knows why!), the writer threw up his hands in despair and refused.

Nevertheless, the lives of each, of Clara and of Harry, involve just one person. The early Harry (say the Boston Harry) and the later Harry (say the Michigan Harry), however different in personality and interests, are one and the same person, as are the early Clara and the later Clara. There was not a series of different persons in the case of Harry any more than there was in the case of Clara. To put it another way, neither Harry nor Clara ceased to exist at any stage of their lives.

Generally speaking, there seems to be little connection between the question of the story, or narrative, unity of a series of events --- of whether the events form a coherent story --- and the question of personal identity over time --- of whether the events are events in the life of one and the same person. There can be personal identity over time without narrative unity. Harry’s life is a clear example. And, as we shall see, there can also be narrative unity (at least of a certain kind) without personal identity.

Many thinkers, both traditional and recent, disagree with these intuitions. They hold that there is a close connection between narratives and personal identity. Their views vary from the extreme that a person’s having a persisting unifying narrative is necessary for personal identity, to the weaker view that having a narrative strengthens the psychological connections essential to personal identity. The thinkers include both Continental philosophers, as well as more analytical philosophers such as McIntyre, Frankfurt, Bratman, Schechtman, and Dennett, to mention just a few.¹

In what follows I shall argue against these views. My paper will be divided into a number of sections. In Part I, I shall say a bit about traditional and recent treatments of personal identity and of stories. In Part II, I shall propose and then rejection a number of hypotheses connecting narratives and personal identity over time. In Part III, I shall look at indirect attempts to defend the connection between narratives and identity over time by way of the idea of a person’s characterization identity, and then show that these arguments also fail. I shall concentrate on personal identity over time, but hopefully much of what I say can be extended to deal with identity at a time, such as multiple-personality cases.

The Idea of Personal Identity and the Idea of a Story

Personal Identity

Let us begin with a few words about personal identity. Philosophers since ancient times (Martin and Baresi, 2006) have been interested in the question of personal identity, which for them has been primarily the question of what it takes for a certain person today to be one and the same person as a person

¹McIntyre, Schechtman and Dennett, as well as Sartre and Ricoeur, are cited and discussed by Galen Strawson (2005, 2007).
Personal Identity and Stories

yesterday or tomorrow. What is it for there to be just one person (numerical identity) rather than two? The question is a subspecies of the general question of identity for physical objects (stones), living things (trees), artifacts (ships), groups (the Sao Paulo Symphony), and even works of art (Villa-Lobos’ Bachianas Brasileira, No. 4). And it is often approached by way of puzzle cases, including Theseus’ ship, Locke’s prince and cobbler, and more recently examples of amnesia, brain transplants, beaming down, and splitting.

There are a number of theories of personal identity. Many agree that personal identity requires the persistence of personhood: at every stage in the life of one and the same person, a person must be present. But this is not sufficient. My person-stage right now and Hillary Clinton’s a second later are persons, but clearly not stages of one person. A popular view is that the supplement involves some psychological connection, some psychological continuity or unity, perhaps underpinned by a physical body or brain. And of course the question then becomes: what kind of psychological connection will do the trick? Answers have ranged from the persistence of detailed psychological features (memories, beliefs, character traits) to the persistence of merely a core psychology (Unger, 1990; Adams, 1992). I think it is fair to say that narrative accounts of personal identity fall under the heading of psychological-connection theories.¹

Stories

What are stories, or narratives? Central features of stories include a perspicuous representation of a temporally ordered series of events (Carroll, 2001), a single subject taken in a broad sense, and various causal linkages between the events. The Peloponnesian War, Hamlet, the classical sonata, and of course Clara herself all count as possible (indeed actual) story subjects. It would be wrong to say that every event in the story subsequent to its start is caused by an earlier event in the story. Pearl Harbor occurs in a history that I am reading of Lyndon Johnson’s political career. Earlier events in the narrative have little, if any, causal bearing on that event. Nevertheless, it is an important event in LBJ’s story insofar as it has significant effects on his future career. Additional features of the events of a story that are often present include being part of a broader development or plan (Sherlock Holmes’ plan to catch Moriarty), instantiating moral generalizations (‘Virtue ought to be rewarded; the guilty deserve to be punished’), and fitting into the traditional literary genres of tragedy, comedy, romance, or satire/irony (White, 1973). The events may also be represented in such a way as to create various emotions in the

¹Two interesting developments in the last few decades are the views of Derek Parfit and Eric Olson. Parfit argues that what matters when we worry about survival is more general than identity, namely certain psychological connections that obtain even in splitting cases (1971). Eric Olson’s view is that personal identity is just animal identity: person is not a substance sortal (Wiggins’ expression), nor perhaps even a sortal at all. Olsen’s view is developed in his 1997 and 2007. Thompson (1997) also seems to hold this view. In what follows I shall assume narrative accounts are relevant to Parfit’s general psychological relation as well as to identity. I shall also have to assume that Olsen’s view is wrong (I must confess I am not completely sure how to show this).
reader or audience, such as hope, suspense, and relief (Velleman, 2003). Stories may have various aesthetic features such as foreshadowing, symmetries of various kinds, and emotional balance.

The idea of a story, or narrative, is admittedly vague. I want to stress at the outset that I am using the idea in such a way that stories need something more than a bare causal sequence of events. I do not want to count the following sequence as a story. At midnight Harry was asleep; at one o’clock Harry remained asleep (partly) because he had been asleep at midnight; and at two o’clock Harry was still asleep because he had been asleep at one. If bare causation were sufficient for narrativity then almost any causal account of personal identity would connect personal identity to having a narrative. Stories are much richer than that. They require at least some of the additional features mentioned above, or features like them.¹

**Are Narratives Needed in an Account of Personal Identity?**

Let us try to come up with a plausible theory linking personal identity (over time) to narratives. I shall begin with (what I shall call) objective theories and then turn to subjective ones.

The simplest objective theory runs as follows:

*The Simple Objective Theory*

Early person $E$ (picked out at an earlier time) is identical to (is one and the same person as) later person $L$ if and only if there is a true and coherent story into which experiences (psychological events, states, etc.) of $E$ and of $L$ can be fitted.

This theory is objective since it does not require that anyone, including the subject, know or tell the story, or have the story in mind. It requires simply that there be such a story. The requirement that the story be true should be obvious, although I shall say more about this later in my talk.

A glaring objection to the simple theory in this form is that it fails as a sufficient condition of personal identity. The early chapters of a biography of Clara describe the musical experiences of her mother as a teenager. These experiences of course occurred long before Clara was born, but they help to explain why her mother was later to push Clara in a musical direction and hence why even later Clara took piano lessons. Her mother’s earlier experiences and Clara’s later piano-lesson experiences can be fitted into a true coherent story. The simple objective theory then requires us to say that Clara and her mother are identical!

We could try to deal with this objection by stipulating that a coherent story must involve only one subject and we would also have to add that the subject must be a single person. This move would be not only ridiculously narrow but also obviously circular.

---

¹Thanks to Fred Adams for getting me to make this point.
If the simple objective theory is to work, then, it can only be as providing a necessary condition for personal identity. So, let us look at that revised account. The revised account is that objective narrative unity is at least necessary for personal identity over time.

Harry’s scattered life seems to present us with a counter example. Harry the biking enthusiast today is clearly identical to Harry the bank teller a few years back, but is there really a coherent story that can be told linking his banking experiences with his biking experiences? Today Harry hardly ever thinks about money and a few years ago he hardly ever thought about biking or sports. Perhaps a more skilled writer than the hired biographer could construct a true and coherent story linking the two, but I am not holding my breath.

Another problem with the simple objective account (even revised) is the many-story problem. There may be more than one story per person; indeed, there are often many. You have heard parts of the story of Clara the pianist, but there is also the story of how Clara became a very good cook. I will not go into details, but what is important is that unlike her romantic involvements Clara has been able to keep her cooking activities somewhat separate from her piano activities.

What is the many-story problem? Well, on the simple objective account Clara playing the piano today may not be identical to the child Clara helping her mother to cook chicken livers, but she may be identical to the child Clara practicing scales. Similarly, on this theory Clara cooking lunch today may be identical to the child Clara helping her mother to cook, but not identical to young Clara practicing her scales. Indeed, it may turn out that even today Clara practicing the piano in the morning is not identical to Clara cooking in the evening. The biography Clara the Pianist does not mention anything about her cooking activities; the biography Clara the Cook does not mention much of her musical life. And all this, even though Clara today remembers in detail both her early experiences of cooking and her early experiences of practicing the piano. The many-story problem forces us to hold that identity is story relative: that the person who figures in the cooking history is not identical to the person who figures in the music history. And that seems crazy.

The Chain Objective Theory

A natural further revision of the narrative approach to personal identity is to replace the requirement of a direct story connection with that of a chain-story connection. Here we are following the lead of memory theorists in the Lockean tradition (who appeal to the ancestral of a direct memory connection). The chain theory is that for one person to be identical to another there need not be a single story that unifies the two, but only an overlapping series of stories. Consider Harry’s college life in Florida and his life teaching English in Morocco. There is a kind of story that can be told about his Florida life: he spent much of his time hunting for the best bars. And there is a pretty unified story that can be told about his Moroccan experience: he really did try to help

---

1Perhaps we could turn it into a sufficient condition by combining the story condition with some kind of bodily or psychological connection.
his students as much as he could. But there is no story, or not much of a story, that links those two periods of his life. Nevertheless, the two stories form a chain: the stories overlap. The end of the Florida story contains an event that also occurs at the beginning of his North African story: recovering from a hangover in the spring of his last year at Florida he received, out of the blue, a long distant telephone call from Casablanca offering him the teaching job.

The chain objective theory, then, is the theory that earlier person E is identical to later person L if and only if there is a chain of stories connecting experiences of E with those of L.

Sadly, at least for narrative supporters, the chain theory fares no better, and in some ways worse, than the simple theory. Like the simple theory, the chain theory fails to provide a sufficient condition of personal identity, unless supplemented by other non-narrative conditions. Secondly, unless the notion of a story is stretched to cover almost any sequence of events, there may well be episodes in a person’s life that are not covered by any story, or, if they are, remain unlinked to relevant stories of early and later periods of the person’s life. Harry’s nervous breakdown after he came back from North Africa was such an episode. During that period his day to day life had very little structure: he seemed to be living from moment to moment. Perhaps, a psychiatrist could construct a story-like account of Harry’s life during this period. The story would describe the waxing and waning of Harry’s depressive symptoms and perhaps be entitled ‘Story of a Depressive Episode.’ But finally, even if we count the psychiatrist’s description as a ‘story,’ it is not at all clear that there was any overlap between this story and stories of Harry’s adjacent past (his Morocco experiences) and his adjacent future (his bank-teller experiences).

Memory theorists of personal identity deal with a similar problem by appeal to potential memory. Last night while I was dreaming I was not having any memories of what happened yesterday afternoon, but I still had potential memories of yesterday afternoon: had I been awakened I would have remembered; or better, while asleep I had various neurological memory traces. Chain narrative theorists cannot make this move however. During Harry’s depression there was no actual story but there was a potential story. What could this possibly mean?

The most serious objection to the chain theory is that it fails to live up to the purpose for which narrative accounts were introduced in the first place. The motivating idea was originally that traditional and contemporary psychological-continuity accounts did not give a life enough unity to achieve personal identity, and narrative connections, at least as a supplement, would remedy that defect. A life, like Harry’s, may have great memory continuity (even in the nervous breakdown period his potential memories were preserved), but at the most only chain-narrative unity. Forget, for the sake of argument, that the chain was broken during Harry’s breakdown. Then the memory connections and continuity in Harry’s life provide much more unity than the chain-story connections. Why do we even need the latter connection for personal identity?

---

1This hope for more unity than psychological continuity can allegedly give is certainly what motivates Schechtman’s narrative account of personal identity in her (1996).
Subjective Theories

Subjective story theories require not simply that there be a true story linking earlier and later experiences of a person but also that the subject have or hold the story: that the story is his or her story. They require, for example, not simply that a possible biographer of Clara’s life could construct a true story linking Clara’s earlier and later piano experiences, but also that Clara herself has or holds that story about herself. Subjective theories can be of the simple or chain variety.

Some work might be done spelling out more carefully what it means to hold or have a story about one’s own life. Ability to tell the story is not necessary. Perhaps all that is required is that there is a kind of tacit or implicit acceptance of the story: it shows up in the thoughts and behavior of the individual.

Unfortunately, subjective story theories of personal identity of any ilk have serious difficulties. They have all the marks against them of the objective theories, since they entail the objective theories. They have the problems of circularity, the many-story problem, and the problems of chain narrative accounts.

There are two additional worries for subjective theories. First, there will be many cases in which the subject herself or himself has no story, even if there is a story. Harry during his Florida days and certainly during his nervous breakdown had no story of his whole life or even of his present life, although perhaps a novelist, like Tom Wolfe, or a Freudian psychiatrist could have constructed one.

Second, subjective theories seem to push us into not simply the story-relativity of identity, but also into the temporal-relativity of identity. Let us see how.

Suppose that today Clara today holds no story of her life that includes an early disastrous romantic episode with Johann. In Freudian terms, she has repressed the episode altogether. She does, however, have a coherent story of most of the rest of her life. Next month, however, after some therapy, she comes to remember and acknowledge the Johann episode as part of her life. On the subjective account we would have to say that today Clara is not identical to the woman of the Johann episode, but that next month she is! In other words, we are pushed into holding that personal identity is temporally relative. And that seems wrong.¹

Characterization Identity and Identity over Time²

At first go around, then, it seems that stories do not play much of a role in

¹Andre’ Gallois has defended a temporally relative identity view in his book on identity (2003) but for quite different reasons involving Theseus’ ship and other examples.

²As will be obvious, my ideas in this section owe much to Velleman (2006) and to Galen Strawson (2005, 2007). I also am indebted here to Rob Noggle (1998) as well as his helpful remarks in conversation.
constituting personal identity over time. Narrative unity of experiences (or other psychological events or states) is neither necessary nor sufficient for personal identity. But perhaps we should not give up so quickly. Let us explore a more indirect, and indeed popular, route to connecting stories and personal identity.

The intermediate stop on our new route involves a different, but still familiar, notion of a person’s identity. This is the notion of what I shall call (borrowing from Maria Schechtman) characterization identity. The kind of identity at issue here is not ‘the relation which every object bears to itself and nothing else, but rather … the set of characteristics that make a person who she is.’ It is the ‘kind of identity that is at issue in an “identity crisis”’ (73–92).

Many different features can be relevant to a person’s characterization identity: his values, his character traits, his self image, his plans, his behavioral quirks, and his various social and professional roles. For Harry Frankfurt (no relation to our Harry!), a person’s characterization identity is partly constituted by ‘contingent volitional necessities,’ motives that the person somehow cannot alter (1999, p.138; cited in Velleman, 2006, p. 334.). Clara’s very strong desire to pursue a piano career is part of her identity in the sense of characterization identity.

We now have three ideas: the idea of personal identity over time, or simply, identity over time; the idea of characterization identity of a person; and finally the idea of narrative unity of experiences (or other psychological states or events). The indirect route tries to connect the first and third ideas by way of the second.

Here is one way the indirect route might go. It consists of four steps (three premises and a conclusion).

1. Having a characterization identity is necessary for being a person (at a particular time).
2. Persistence of characterization identity is necessary for identity over time.
3. Having a narrative of one’s life is necessary for having a characterization identity.
4. Therefore, having a persisting narrative of one’s life (or narrative continuity) is necessary for identity over time.¹

Most of the philosophers mentioned at the start of this talk sympathize, or at least have sympathized, with one or more of these steps.

Frankfurt, for example (op. cit.), accepts premises 1 and 2. Remember that for him characterization identity is partly defined in terms of motivational necessities. These motives constrain what one intends to do, but are somehow unalterable. We might think of Martin Luther’s motives in saying ‘I cannot …

¹This version of the indirect route is a bit strong if the second premise is taken to require the continuity of a single characterization identity. Perhaps we can weaken it by allowing for gradual change in characterization identity.
recant’ (or more popularly, ‘Here I stand. I can do no other’). Frankfurt’s example is Agamemnon.

For Frankfurt, Agamemnon at Aulis experiences a conflict between ‘two equally defining elements in his own nature: his love for his daughter and his love for the army he commands. … When he is forced to sacrifice one of these … [his] volitional unity … [becomes] irreparably ruptured … [and] there is a sense in which the person he has been no longer exists’ (ibid., p.139; quoted in Velleman, 2006, p.335).

Does this indirect route succeed in getting us from personal identity (over time) to narrative unity? Sadly, at least for story-theory fans, it does not. There are obstacles at each step of the way. Here are some.

To begin with, like the term ‘self’ the term ‘identity’ in the characterization sense (or alleged characterization sense) is really much too unclear and vague (see Olson, 1999). I have a friend who leads quite a different life today than twenty years ago: she was formerly a fashion model in Los Angeles and is now a writer in New York. Her projects, general motives, and daily habits have changed dramatically. Still, her psychological chemistry or style --- her gestures, idiosyncratic vocabulary, and patterns of associations --- remain remarkably the same. Has her characterization identity changed? I can’t say. Again, I myself have many ‘roles’ today: teacher, friend, bar-piano player, mountain hiker, brother, son, and so on. Some of these roles are closely connected, others very little. Are all these roles part of my identity in the characterization sense, or is one of them paramount? Again, I can’t say. One thing I forgot to mention about Harry is that almost all his life he has had hypochondriac thoughts that seem very much a part of him, although he says that he would like to get rid of them. Are these tendencies part of his identity in the characterization sense? Once again, it is not clear.

Still let us assume that there is some useful concept of characterization identity that we can work with. But then it is easy to criticize all three premises.

**Criticism of Premise 1**

Is having a characterization identity really necessary for being a person at all? Pick your favorite definition of characterization identity --- in terms of motivational necessities, or of dominant social role, or psychological chemistry. It is not difficult to think of persons, or at least of persons during a certain period of their lives, that lack characterization identities. Galen Strawson has an apt quote from the 20th-century Welch writer, Goronwy Rees:

> For as long as I can remember [write Rees] it has always surprised and slightly bewildered me that other people should take it so much for granted that they each possess what is usually called a ‘character’; that is to say, a personality with its own continuous history…. I have never been able to find anything of that sort in myself…. As a child … I was content with a private fantasy of my own in which I figured as Mr. Nobody (Strawson, 2005; I cannot
locate the quote’s original source).

And, we could add, there is no reason to think that others, however psychologically astute, would have been able to find a personality or character in Rees, either.

**Criticism of Premise 2**

Even if we accept the first premise that personhood requires characterization identity, it is very implausible to hold that persistence of character identity is necessary for personal identity over time.

Take Agamemnon. Let us accept with Frankfurt that Agamemnon’s characterization identity has been deeply changed as a result of his sacrificing his daughter. Still, the Greek tragedies about Agamemnon are surely all about one and the same person. It would be absurd to say that Agamemnon before Aulis and the Agamemnon that returns from Troy only to be murdered are two different persons! The tragedies would lose much of their force (Velleman, 2006, p. 340). Characterization identity has changed, but not personal identity.

Other examples are many. Saint Paul is still the same person after the road to Damascus. Harry is still Harry after his recent marriage and change of lifestyle. I would argue that even Ansel Bourne, the 19th-century amnesiac described by William James, remains the same person after his fugue episode begins (Wilkes, pp. 103-106; Fuller, p.151).

**Criticism of Premises 3**

Narratives are not necessary for characterization identity. Having a characterization identity does not entail having a narrative of one’s life (nor does it entail that there is a narrative of one’s life). Consider Harry now: he is financially secure because of his marriage, he has interests in biking, watching the news, going out to dinner, and other simple daily pleasures; he enjoys being with his wife; he has become kind and generous. However, he really lives very much in the present. He has few, if any, long term projects, except perhaps to continue what he is doing now. He also does not even try to knit his past live into his present one.

I guess that we can say that he has a characterization identity: he has various and relatively stable and unified psychological dispositions. But he certainly does not have a narrative of his whole life, or even of this new period of his life. Nor are we outsiders in assigning a characterization identity to the newly wedded Harry committed to telling a coherent story about his life. Of course, Harry can describe some of his daily actions in terms of short term goals: going to the store to pick up some fresh fish for the party tomorrow, repairing his flat tire because he plans to go on a long bike ride later in the day, and so on. And so can we. But this would be having a narrative, or there being a narrative, only in a trivial sense (Strawson, 2005).
Final Criticism of the Indirect Route: Narrative Persistence without Narrative Unity

The indirect route to the conclusion that stories are necessary for personal identity fails. There is, moreover, an overlooked reason that this strategy fails. The conclusion was really supposed to be that narrative unity is necessary for personal identity. But the conclusion of the indirect route was only that a person’s having a persistent narrative is necessary for personal identity.

The trouble is that having a persistent narrative, or there being narrative continuity, is not the same as narrative unity. There are some people, actually very nice people, who for the last thirty years or so have had a persistent narrative of their lives: their narratives could each be entitled Story of a Successful Academic. Their stories, however, do not unify the events of their lives because the stories are mostly false! For example, the stories emphasize that it was their hard work that led to their success; as a matter of fact, they did not work very hard, they were not very successful, and what they took to be success was mostly luck!

We might try to bring narrative persistence and narrative unity closer together by requiring that a person’s having a narrative requires that most of the narrative be true (I am not sure of exactly how to do this without stipulation). Even if we succeed here, the indirect route to link stories with personal identity will fail because of our earlier criticisms.

Brief Conclusion

Narratives are neither sufficient nor necessary for personal identity. Both direct as well as indirect routes in support of a connection between personal identity and stories fail.¹

I have said little about the problem of identity at a time, the problem of how many persons there are in one body at one time? What should we say about cases of fugues, psychological disassociation, and what used to be called multiple personality? I am pretty sure that narratives will be of no more help with that problem than with the problem of identity over time.

Finally, there are other questions about narratives and persons that are important. There are questions about whether narratives do play a role in people’s lives and whether this is, or would be, a good thing (Strawson, 2005, 2007). These questions, however, are distinct from my question here about stories and personal identity.

References


¹I would also argue against the weaker view that narratives at least strengthen connections that constitute identity. Sometimes they do, but other times they weaken the identity connections.


CHAPTER TWENTY TWO

Why there is 'No Such Thing as Language': Chomsky's Mistake

Patricia Hanna

I examine the Chomskyan view that language is a natural object operating according to specific rules or “laws of nature”, and that it should be studied as such. I argue that this view, despite all its apparent success, is fundamentally mistaken because it cannot handle semantics, in the ordinary sense of a theory of meaning, understood as the bearer of content.

I focus on Chomsky’s discussion of meaning, specifically what he calls “referential theories” of meaning, which include both direct reference (Mill and Kripke) and Frege-Russell descriptive theories. Chomsky’s rejects these because they base meaning on something mind-external. He replaces this with a mind-internal ground—viz., our genetic endowment.

I show that Chomsky’s criticisms misidentify the source of their failure. The referentialist’s mistake does not turn on the distinction between the mind-independent and the mind-external, rather, it lies in the difference between the mind-independent and the mind-dependent. Genetic endowment is as mind-independent as the so-called “external or real world”, hence Chomsky fails to provide any actual advance on the theories he rejects. The ultimate irony in this is that Chomsky’s fundamental mistake lies in a petitio which shares its form with the petitio he accused B. F. Skinner of committing over 50 years ago (see Chomsky 1959).

Definitions

Chomskyan linguistics uses terms with technical meanings. To avoid confusion, here is a list of the main terms.

1. Generate means “assign a structural description to”. It does not mean “produce” in the sense of creating the actual bits of language we use in real-world discourse.

---

1I use Chomsky as the main source of the views discussed here. While various linguists and philosophers, including Chomsky himself, who identify themselves as “Chomskyan” may not accept all of them together, they are all part of various versions of the Chomskyan program.

2The character of the rules is open to interpretation: they range from universal invariant laws to principles and parameters. The specific understanding of “laws of nature” does not affect the arguments I offer here.
A grammar generates language if it provides a complete structural description of all possible linguistic output. Since this output is infinite, a grammar must assign a structural description to an infinite set of outputs.

2. Competence and Performance

(a) Competence is the speaker-hearer’s (s/h) knowledge of his language. It should not be identified with what any real-world s/h actually knows or claims to know. Competence is defined relative to an ideal s/h, someone who has no memory limitations or lapses in attention, who is never tired, forgetful or at a loss for words, and so on. Hence, competence is the ideal speaker-hearer’s knowledge of language. The complete definition is of competence is “the theory or set of rules which an ideal s/h must know in order to account for the generation (see 1 for the sense of “generate”) of linguistic output”.

(b) Performance is the real-world s/h’s actual use of his language in concrete situations. While competence is never actualized in the real world, performance has no existence apart from the real world.

Competence is perfect, performance is riddled with errors, false starts, &c. A theory of competence accounts for the assignment of structural descriptions, with no attention to who’s doing the assignment or how things might go wrong. A complete theory of performance must take account of memory limitations, inattention, motivation, context, setting, and so on.

3. A grammar is a set of rules which assigns structural descriptions to all the sentences and phrases of the language under consideration.

4. The Universal Grammar (UG) is the set of rules which underlies all humanly possible languages.

5. The Faculty of Language (FL) is whatever set of systems humans have which accounts for language. In Hauser et alia 2002, a distinction is drawn between

…the faculty of language in the broad sense (FLB) and in the narrow sense (FLN). FLB includes a sensory-motor system, a conceptual-intentional system, and the computational mechanisms of recursion, providing the capacity to generate an infinite range of expressions from a finite set of elements. We hypothesize that the FLN only includes recursion and is the only uniquely human component of the faculty of language (1569).

6. I-language and E-language are amongst the most troublesome concepts in the literature. One is tempted to think of them as
follows.

(a) **I-language**: internalized theory of language (or grammar) which any individual s/h must know in order to be credited with knowing language. In “Internalist Explorations”, Chomsky says that *I* is used “…to suggest internal, and individual, and also intentional…” (Chomsky 2000c, 169).1

(b) **E-language**: externalized language independent of the properties of the mind/brain. This is closer to Saussure’s concept of *langue*, a socially constructed and maintained set of rules governing a specific natural language. (See Chomsky 1986, 19–21.)

Both I-language and E-language provide “norms” for the s/h (see Chomsky 2000b, 72, and Chomsky 1986, 19–21). How do these norms function under biolinguistics?

I-language (internal language) [is] a state of the computational system of the mind/brain that generates structured expressions, each of which can be taken to be a set of instructions for the *interfaces* systems within which the faculty of language is embedded (Chomsky 2007, 1).2

Consequently, UG is only one of many I-languages, but it is the only one that is innate, i.e., not learned from or based on experience. Hence, speakers have an I-language for every language they know. The definition of *E-language* is not affected by the shift to the biolinguistic.

**Explaining Language**

Chomsky argues that language is associated to a physical system; as Eric Lenneberg says, it “has a biological foundation”. This does not commit him to holding that the study of language is exhausted by an analysis of this physical system. Language is more than a physical entity; in addition to its physical properties, it has properties of a sort that the digestive system, say, cannot have. This is Chomsky’s mentalism, and it explains why he uses the term “mind/brain”, but not “mind/stomach”.

Chomsky’s view that language is a natural object has two functions.

1. It makes clear that language is not a social phenomenon, it is not the product of human invention, in the way the stock market is. Thus, when he says that language is *real*, he does so to contrast it

---

1 Aniko Csirmaz suggests that “idealized” might be added to the list of I-terms. This would have the advantage of making more perspicuous I-language’s status as an abstraction and of showing its role in the sort of scientific inquiry Chomsky is interested in.

2 These interface systems are, roughly, the mental and the sensorimotor.
with things which are mere conventions or inventions of the human mind, all of which have an inherent element of arbitrariness.

2. It allows Chomskyans to regard the study of language as a natural science. This entails that every aspect of our study of language must answer to the criteria of the natural sciences, and we must use the best methodologies of these sciences. Consequently, language is to be understood and explained through the creation of models answering to the standards of formalization, elegance, simplification and mathematics, in just the way the world is understood and explained by physicists.

These fit with understanding UG as a “theory of the genetic endowment for languages”. Since language is biological, the theories—i.e., grammars—are theories about biological systems.

This approach, or some variant of it, has expanded our understanding of language. It has proven especially useful in syntax and phonology. I don’t claim to know the complete story of this success, but something along the following lines explains part of it.

- Syntax lends itself to the formal, axiomatic approach of idealized modeling and theory building. If the job at hand is generating structural descriptions, it is easy to take the formulation of a set of rules adhering to the strictures of mathematics/formal logic as the aim of your research. In addition, there is a further advantage: the linguist can use biological features and theories to help eliminate competing, but equivalent, theories, thus answering (or at least mitigating) Quine’s concerns about the underdetermination of theory by data.¹
- Phonology also fits. The sounds we can make and hear, and the ways in which we can combine them are deeply physical. For example, the human auditory system simply cannot process sounds above a certain frequency; this is a physical limitations; hence, focusing on biology allows us to explain many features of phonology.²

Semantics, however, is the odd area out. Even Chomsky senses this.

As for semantics, insofar as we understand language use, the argument for a reference-based semantics …seems to me weak. It is possible that natural language has only syntax and pragmatics; it has a “semantics” only in the sense of “the study of how this instrument, whose formal structure and potentialities of expression are the subject of syntactic investigation, is actually put to use in

¹I would argue that this does not actually address Quine’s criticisms, but I don’t discuss that here.
²In saying this, I am not saying that phonology is exhausted by the study of the organs which produce and process sounds.
a speech community” …There will be no provision for what Scott Soames calls “the central semantic fact about language, …that it is used to represent the world,” because it is not assumed that language is used to represent the world in the intended sense (Chomsky 2000e, 133).

He goes on to claim that his earliest formulation of generative grammar was “influenced by Wittgenstein, Austin and others”. In §6, I show that there can be no such influence. Treating language as a natural object, not a social construct, precludes using any social/conventional aspects of human language to explain meaning. But, the social/conventional nature of language is precisely what accounts for the unique contributions of Wittgenstein and Austin to our understanding of language and meaning.

In the next section I consider why semantics, understood as a theory of linguistic interpretation and content—meaning—defies the Chomskyan biolinguistic approach.

The Semantics Stands Alone

The best current explanation of how language is recursively generated is through Merge, an operation that takes objects already constructed, and reconstructs a new object from them, generating a ‘language of thought’, perhaps in a manner close to optimal (relying on principles [that are not specific to the faculty of language]), with externalization (hence communication) a secondary process (Chomsky 2007, 1).

Describing it as a “language of thought” allows Chomsky to distinguish one sense of “language”—I-language—located entirely inside the s/h, from an outer, public language used for intra-personal communication—E-language. Hence, there are at least two points of interface between language and other systems.

…the systems of thought that use linguistic expressions for reasoning, interpretation, organizing action, and other mental acts; and the sensorimotor systems that externalize expressions in production and construct them from sensory data in perception (Chomsky 2007, 1).

The first is for “the language of thought”; the second, the externalized language of communication, which I call “the language of conversation”.

Turning to the representations, there are two ways in which they are to be assigned interpretations/meanings.

1. I-language: If I am concerned with my representations of my thoughts, and I use language to manage them, as described above,
there is an important sense in which I needn’t worry about the specific semantic interpretation assigned to “sentences”. If I am consistent and I always (or almost always) invoke the interpreted representations in the same circumstances, i.e., in connection with the same sections of my environment and/or my biological-space, everything is fine.\(^1\) For example, I represent my present state as (1) “My leg hurts” and interpret this as (2) I have a good feeling about this argument. So long as I am consistent and do this only when I am (3) in some specific psycho-physio-state—which might be that state we think of as enjoying the taste of a coconut cup-cake—everything is fine. A set of elements in my genetic endowment, could work for this purpose. But, we must be clear: apart from being connected as just described, the interpretations and the representations are not connected by anything that might count as content.

**E-language:** When I-language is externalized to function as a language of conversation, we need more than consistency and repeatability. For communication, we need to explain the *content* we convey to others via language. The content must be sufficiently rich to allow us to distinguish between cases in which it applies and cases in which it doesn’t apply: to put it another way, it allows us to make statements which are true or false, appropriate or inappropriate, and so on. I call this “assertoric content”. At a minimum, a semantic interpretation must account for assertoric content. Hence, in constructing semantic theories, we must deal with an issue which is at least akin to what Scott Soames identifies as “the central semantic fact about language”. In communication, it’s not simply that I want you to arrive at the *same* interpretation every time I utter “A”—of course I want that—I want you to be consistent with respect to a *specific* interpretation or range of interpretations: that is, I want you consistently to assign “A” same interpretation, and I want it to be the one I associate to it. In order to achieve this, we need access to full-blooded concepts with assertoric content.

The problem for communicating our thoughts, as contrasted with organizing them, is that intuitively we want there to be some connection between the sentence, the interpretation/meaning, and the world.

One way thought to solve this problem is to connect semantic interpretations to something mind-independent, either directly, as a referential theory of meaning does, or indirectly through properties or denotation, as a descriptive theory does. Chomsky correctly sees that these theories do not give

\(^1\)Chomsky rejects any account of meaning which ties it to the external world, as the referential theories would have it. In saying that we want to stick to the “same sections of my environment and/or biological-space” I intend to honor this restriction: the environment at issue *just is* the interior environment of my mind/brain.
language assertoric content. Unfortunately, he draws the wrong conclusion from this. Focusing on the referentialist’s appeal to a mind-external world leads Chomsky to try to give meaning a mind-internal grounding. But this merely pushes meaning inside the s/h, with the result that semantic interpretations are formal and empty of content.

Of course, Chomsky would reject this reading. The next section gives a more detailed discussion of Chomsky’s attempt to ground meaning in something internal to the mind/brain.

**Chomsky on Semantic Interpretation**

Locke recognized that persons are individuated in part by psychic continuity. Thus a single body could house two distinct persons if two distinct psychic continuities were connected to it, and a person could persist even if its substance became completely different, properties of our concepts exploited by science fiction. These are all properties of the common notion or innate idea of person. Locke argued further that ‘person’, unlike ‘tree’ or ‘cat’, is a ‘forensic term, appropriating actions and their merit’, belonging only to ‘intelligent agents, capable of a law, and happiness, and misery’, and hence even more remote from mind-independent physical investigation, and at the heart of our moral faculty and intuitions (Locke, 1689/1975: p. 346) (Chomsky 2007, 9).

To understand ‘person’, we must look beyond the mind-independent world and our observations of it, and “…focus attention on intricate aspects of the world by resort to our cognoscitive powers” (Ibid., 9). These powers give us access to precisely those features of the concept person not accessible by observation of the world.

Generalizing this, it follows that the language of conversation depends on our having access to representations given by our shared “cognoscitive powers” in terms of concepts which are cannot be “read off” nature. These concepts form the basis of the assertoric content in meaning. Since this content is not present in nature (the world), it must be innate in the sense that any features of I-language which are not accessible experientially are said to be innate. This rich stock of semantic information, which I call “semantic markers”, allows us to augment experience and build concepts. Thus, communication cannot rest only upon the mind-independent world, but must involve our shared internal structure, including “cognoscitive powers” and semantic markers.¹

Elsewhere (Hanna and Harrison 2004), I argue that the referential approach

¹This claim is not new. John Searle discusses it. (See Searle 1972 and 2002.) What is interesting is that the Chomskyan program is steadfast in its refusal to comes fully to grips with the the sorts of issues raised by Searle.
to meaning holds that assertoric content is part of the world, and hence meanings are real. I call these “referential realist” theories of meaning. Chomsky’s criticisms go some way to showing why referential realism does not succeed: the world doesn’t have assertoric content, it is something which only humans can bring to the world. However, Chomsky’s proposal that genetically given semantic markers provide access to assertoric content commits the same mistake as the referential realist. Human biology, in the form of shared “cognoscitive powers” and “semantic markers”, removes meaning from the mind-external world, and puts it into a mind-internal world. But the mind-internal world is itself mind-independent: we no more invent our genetic make-up than we invent the Grand Canyon. Consequently, it is incapable of giving language assertoric content.

Against referential realism we develop a view of meaning which connects language to the world through the social practices which are invented and maintained by us for the purposes of communication. In the next two sections, I show that Chomsky’s solution and his appeal to Wittgenstein and Austin amount to an attempt to access these social constructs. At the point where he seems to have saved his theory from vacuity, he assumes that meaning rests on shared practices and conventions, thus begging the question in much the way he accused Skinner of having done (see Chomsky 1959).

**Meaning and Content**

Communication is not a matter of producing some mind-external entity that the hearer picks out of the world, the way a natural scientist could. Rather, communication is a more-or-less affair, in which the speaker produces external events and hearers seek to match them as best they can to their own internal resources. Words and concepts appear to be similar in this regard, even the simplest of them. Communication relies on largely shared cognoscitive powers, and succeeds insofar as similar mental constructs, background, concerns, and presuppositions allow for similar perspectives to be reached. If that is true—and the evidence seems overwhelming—then natural language diverges sharply in these elementary respects from animal communication, which appears to rely on a one-one relation between mind/brain processes and “an aspect of the environment to which these processes adapt the animal's behavior”, to quote cognitive neuroscientist C. R. Gallistel's introduction to a volume of papers on animal communication (Chomsky 2007, 10).

Suppose I want to tell you that I feel good about my arguments. Since I have access to language, I say “I feel good about my arguments”. The external event is my making these sounds. You hear them, take them as input for your FL and “seek to match them as best you can to your own internal resources”. In a normal case, we not only have our shared cognoscitive powers and innate
semantic markers, but we also a wide range of experiences, vocabularies, concerns, backgrounds and presuppositions in common. All this aids you in your construction: Success! But what does this success consist in?

The problem is not that it’s absurd to suppose that we have shared cognoscitive powers, surely we do else we would not have evolved as we have. It’s not even that these shared powers aren’t crucial, even necessary, to the development of language, whether it’s used to organize one’s own thoughts or to send messages across continents to one’s lover. The problem is that all they can possibly account for are our sensori-motor interactions with an environment, whether it’s internal or external; alone, this gives no explanation of how language mediates our ability to communicate with others. For this, we need to supply the semantic interpretations with content, which is what semantic markers are supposed to do. The question, however, is how this is supposed to happen.

Chomsky explicitly rejects referential theories of meaning, hence the world doesn’t explain it. Interaction with the world is the business of E-language, but conceptual meaning is a prerequisite of this interaction, hence it cannot arise from E-language. Therefore, conceptual meaning must reside at the level of I-language—in our genetic structure. This is clear from Chomsky; however, we are still owed an explanation of how it works.

Consider the startle response psychologists observe when infants are given sensory input that doesn’t correspond to how objects “normally” act: say an object’s emerging with an entirely different shape to the one seen going behind a screen. It seems plausible to hold that if an infant didn’t have this reaction, we could not teach it about physical objects, their continuity, &c. However, having said this, we can still ask when and how the child acquires these concepts, as contrasted with reacting to the passing show. Identifying a kind of attention and reaction to her surroundings which allow us to teach her these concepts, is not identifying what constitutes either acquisition or understanding of the concepts.

We may discover species-wide psychological/biological features which are required for acquiring and understanding concepts. But, by itself, the most this could yield is a pairing of sounds with something in the “environment”, something of the sort that occurs in what P. F. Strawson called “feature placing language” (Strawson 1964). Wittgenstein argued (Wittgenstein 1968) that ostension alone—pointing and saying—can only produce this sort of association between linguistic items and the world: it can only produce “labels”, not names or concepts.

Feature placement or labeling does not have the conceptual weight needed for concept acquisition. For this we need assertoric content: this constitutes the step from label to meaning. Only when we have meanings can language represent our thoughts, whether internally for our own purposes or for communication with one’s fellow beings. It is obvious that we need this for the language of conversation; however, it is equally true of the language of thought. If we exempt I-language from this requirement, two problems arise.
1. The semantic “interpretations” will be as devoid of content as the logic of the *Tractatus* (Wittgenstein 1961).

2. Since the language of thought cannot account for the assertoric content of the language of conversation, it is impossible to explain how communication is possible.

These problems are the same: as language cannot be the bearer of content, it is forever doomed to being “about nothing”.

Chomsky wouldn’t agree. He holds that semantic markers, along with genetic endowment, shared cognositive powers and propensities to react and attend to our environment, produces assertoric content. On one level this seems right, but I argue that it is mistaken. Why does he seem to have succeeded?

Consider this: “Communication relies on largely shared cognositive powers, and succeeds insofar as similar mental constructs, background, concerns, and presuppositions allow for similar perspectives to be reached”. Chomsky tosses this out as if it were nothing very special. In point of fact it is very special, and it is inconsistent with his contention that language, including semantics, is a natural object.

By claiming that “similar mental constructs, background, concerns, and presuppositions allow for similar perspectives to be reached”, Chomsky calls in to play all the forces of our shared human world, the world that is constructed and organized according to our decisions, our conventional ways of acting, science, and so on. It precisely in virtue of this social structure that he seems to have saved meaning and connected language to the world. But begs the question: it relies on a social basis for meaning, which is inconsistent with his claim that language is a natural object.

In the next section, I give a sketch of what teaching a child a concept might actually consist in, and show that this is not dependent on anything like semantic markers as used by Chomsky.

**Concept Acquisition**

Acquiring a concept involves more than acquiring a word. It requires more than being able to apply the word correctly. It requires knowing when to deny application of the word: knowing when to say “No, not C”. And this isn’t all, it requires knowing when to withhold the word altogether: when to say “No, not C, but not not-C either”.

Imagine a three step process. First, a child learns that tomatoes and raspberries and flashing lights on a police car are red, and she always says “Red” when she sees one of these. But then one day, she encounters a yellow tomato and a black raspberry and a flashing blue light on a police car. She might continue to say “red” in these contexts (in which case we would realize that what she has identified with “red” is not the color, but the fruits and the lights on police cars), and we would teach her that these are “not red”. This is
the second step. Finally, third, we continue the process of teaching, and start to ask her about other things: “Those ideas of yours, are they red?” And here she needs to say: “Mommy, ideas aren’t red”. So Mommy asks, “So if are they not red, then, are they blue?” And here the right answer is, “Mommy, ideas aren’t red or blue or green or yellow or any color, they just aren’t colored”.

To be credited with the concept, the child must learn what Wittgenstein calls a word’s “post in the language”, its logic, how it relates to other concepts and what we can and cannot do with it. We must learn its categories: not just its syntactic category (noun, determiner, &c), but its location in conceptual/logical space. We must learn that it’s a color term, concrete object, number, and so on.

Of course as she acquires the concept red, we can introduce the child to the concept of color, and once she has this, learning new colors will be easier. But, and this is the crucial point against positing innate semantic markers, in teaching her the concept red, we don’t appeal to these abstract concepts or universals. From this, it follows that we do not need to posit them as a part of I-language to explain concept acquisition. Doctrinaire commitments aside, we all know that colors are as readily available for us to look at as cars, raspberries and tomatoes are; it’s through use of these things that we access the concepts of red, color, &c.

The same applies to that concept which so impressed Chomsky: person. Learning what a person is doesn’t require that one already understand ‘psychic continuity’ or ‘law’ or any of the other abstract and socially embedded concepts cited by Chomsky. We learn what persons are at the same time we learn what moral agents are, what following a law consists in, what we mean when we talk about psychic continuity. None of this requires that we treat any concept or set of concepts in the group as primary, as Chomsky presupposes.

Conclusion

If I am right, the semantic interpretations Chomsky is entitled to generate from language as a natural object—a “genetic endowment”—are content-free: semantics without meaning. Any language equipped only with such a semantics, is not a language at all.

References

In his influential article, “How is Weakness of the Will Possible?”, Donald Davidson proposes the following principle:

P1. If an agent wants to x more than he wants to y and he believes himself free to either x or y, then he will intentionally x if he either x-s or y-s intentionally.\(^3\)

Though Davidson admits that the ‘critical words or phrases’ in his statement of the principle are ambiguous and may be interpreted in ways that make it turn out false, he nevertheless holds that there is a pertinent interpretation in which P1 is self-evidently true.\(^4\) This is accepted by many philosophers of action, who regard the principle as an essential component of a theory of how actions are to be explained. Such acceptance is by no means unanimous, however. Perhaps the most prominent dissenter is John Searle, who rejects both the principle and the theory of which it is considered an essential component. His criticism occurs in Rationality in Action, where as part of his general project of ‘weaning people off’ the standard model of action-explanation, he argues that P1 is false both on direct intuitive grounds and on the grounds that it contradicts free will. This criticism is likely to appeal to many philosophers, in particular those who share Searle’s libertarian view of free will. However, I believe that it is based on a misunderstanding. My aim in this paper is to defend P1 against Searle’s criticism by elucidating it in a way that makes it clearer why it is compatible with the intuition that underlies the main objection to it.

---

1A version of this article was first published as ‘Freedom and Wanting More: A Defense of Davidson’s Principle’ in Cheolhak 105: 101-119. I thank the editors of Cheolhak for granting permission that the content of that article be re-published here.

2The title derives from that of the article ‘Man’s Will—Free yet Bound’ by Walter J. Chantry.


Senses of Wanting

Before I begin my defense, it is necessary to deal with some minor objections to P1. Recall that Davidson admitted that there were interpretations of the terms involved that made the principle turn out false. However, he regarded objections based on such interpretations as trivial. The following all fall under this category.

At first sight P1 may seem obviously false because, for example, we frequently find ourselves wanting to x more than to y but then y-ing out of duty. However, this is to interpret wanting in a narrow sense, and not the one that Davidson intends. Elsewhere he uses the phrase ‘pro attitude’ to capture what he has in mind, which includes any attitude that ‘directs’ an agent towards a certain action. As Mele more aptly puts it, ‘to want to A is to have some motivation to A.’ Thus one who is motivated by duty to y wants to y in the relevant sense of ‘wanting’.

A second objection is that P1 is false because, for example, I want more to be kind to my friend than cruel, yet on occasions I am nevertheless cruel to her. Here we need to distinguish between occurrent and standing wants. An occurrent want has to do with my current motivational state, whilst a standing want is a disposition to have an occurrent want that manifests under certain conditions (say, when I am in control of myself). In the case above, I am disposed to want to be kind to my friend, though this disposition need not always be manifested: there are times when I may want, in a moment of incontinent spite, to be cruel to my friend. On such occasions I still have the standing want not to be cruel, though I occurrently want to be cruel. The ‘wanting’ of P1 is occurrent.

A third objection is that P1 is false because, for example, I may want to buy a watch more than not, but nevertheless choose not to buy it on account of the expense. What is going on in this example is that I want to buy the watch more than not in respect of whatever it is that makes it attractive to me (say, its beauty), but I want not to buy it more in respect of its expense. It could be that I choose not to buy the watch because the latter respect has the upper hand in my mind. That is to say, overall I want not to buy the watch more than to buy it. It is this ‘overall’ sense of wanting that is intended in P1.

Of course, it could also be that neither respect has the upper hand in my mind. This would be a case of a motivational tie, that is, one where overall I do not want one option more than the other. Here P1 is silent—it makes no claims about how agents act in the case of a motivational tie.

It should be noted that talk of an ‘overall’ sense of wanting does not commit one to the view that all wants can be compared on a single scale. For example, suppose that I want to x more than to y in some moral respect, but want to y more than to x in some non-moral respect. To say that overall I want to y more than to x need mean nothing more than that I have lost sight, to some degree, of

---

1Davidson (2001), pp. 3-4.
3Cf. Ibid.
what I want in the moral respect. It need not mean that the non-moral motive is "stronger", on some single-scale comparison, than the moral motive. If I do not lose sight of what I want in the moral respect, then the situation could be that of a motivational tie—that is, one in which overall I do not want one option more than the other. As just observed, this is a scenario about which P1 is silent.

An Argument for P1

I will begin with a positive argument for P1, which I will refer back to later. This argument is meant to be highly suggestive, if not conclusive. Consider the following example. Bill is offered a choice between a gift of $100 and one of $1, with no strings attached. He chooses the $1. What would one’s reaction be to such an event? Most people would likely be puzzled by it. It would be natural to ask, ‘Why would he want to take $1 more than $100?’ There are two things to note about this. The first is the puzzlement. It is of course reasonable to suppose that an agent would want to take $100 more than $1. But if one’s expectations were not in line with P1, why would one be puzzled when the agent nevertheless chooses $1? The second thing to note is the natural follow-up question. Why suppose that Bill wanted to take $1 more than $100? If one’s expectations were not in line with P1, then why not suppose that the agent wanted $100 more than $1 but simply acted contrary to his preference? Yet the latter explanation, it seems to me, would never be considered in such a case. An observer would always seek to understand the choice in terms of the agent wanting to take $1 more than $100 (or at least being in a motivational tie). In other words, this example indicates that P1 is written into our commonsense psychological expectations. We expect people to act according to what they want more, and if they act contrary to what we figured they would want more, we assume that this is because we were wrong in so figuring. We are willing to revise our suppositions about what the agent wanted more when faced by contrary behavioral evidence even when those suppositions are very firm ones.

One response to the above example would be to say that our commonsense psychological expectations are simply false. But this misses the point. P1 is meant as a claim about commonsense psychology. It is proposed by Davidson as a principle because it is meant to capture a constraint on our commonsense psychological explanations of action and choice, and the above example suggests that it does exactly that. The point is not just that we presuppose P1 whenever we actively seek to explain an action or choice, but that this presupposition is constitutive of the meaning of the key terms within it.
Searle's Objection

Searle's rejection of P1 is part of a broader critique of Davidson's account of the problem of weakness of the will. In "How is Weakness of the Will Possible?", in addition to P1, Davidson presents two other purportedly self-evident principles:

- P2. If an agent judges that it would be better to x than to y, then he wants to x more than he wants to y.
- P3. There are incontinent actions.

His definition of incontinence is as follows:

D. In y-ing an agent acts incontinently if and only if: (a) the agent y-s intentionally; (b) the agent believes there is an alternative action x open to him; and (c) the agent judges that, all things considered, it would be better to x than to y.¹

The problem is clear. Taken together, P1 and P2 imply that if an agent judges that it would be better to x than to y, and believes himself free to either x or y, then he will intentionally x, if he either x-s or y-s intentionally. But this seems to imply that incontinence, as Davidson defines it, is impossible, hence to contradict P3.

Davidson’s solution to the problem is to distinguish between two different logical forms that an evaluative judgement can take. One is that in which the judgement is conditional on a given set of reasons, that is, is of the form that reasons speak more in favour of one action than the other. The other is that in which the judgement is unconditional, that is, is of the form that one action is better than the other tout court. Davidson’s solution to the problem of weakness of the will is that the judgement in the definition of incontinence is conditional, whereas the judgement in P2 is unconditional. So it does not follow from ‘all things considered, x-ing is better than y-ing’ that the agent wants to x more than to y. Hence there is no contradiction.

It is this solution that bears the brunt of Searle’s attack. As he puts it, on a natural interpretation of incontinent action, ‘the thesis is that there are acts such that the agent judges unconditionally it would be better to do x than to do y, believes that he is able to do either, and yet intentionally does y rather than x’ (Searle, 2001, p.224, my emphasis). That is, it is if anything essential to incontinence, or at least typical of it, that the agent acts contrary to an unconditional judgement.

It is at this point that Searle makes his objection to P1. Following on from his criticism of Davidson's solution to the problem of weakness of the will, he asserts the following:

¹Davidson (2001), pp. 22-23. I have swapped the x’s and y’s in D for ease of comprehension.
What the whole discussion shows is that the conjunction of (P1) and (P2) is false. It is not the case that everything that one judges to be the best to do, one really wants to do, and it is not the case that when you have made up your mind and you really want to do something, that you will therefore necessarily do it. (p.228, my emphasis)

The last statement is meant to be the conclusion that P1 is false. Thus his assertion is that his criticism of Davidson's solution to the problem of weakness of the will entails the falsity of P1.

It is, however, difficult to see which part of Searle's discussion shows that both P2 and P1 are false. The discussion is mainly taken up with his argument that one can act contrary to an unconditional judgement. On this I concur. However, this does not show that both P2 and P1 are false, only that at least one of them is. I hold that P2 is false. Searle also thinks that P2 is false. So why then does he infer that P1 is also false?

As stated above, what Searle believes is that it can happen that an agent judges unconditionally that x-ing is better than y-ing and wants to x more than to y in line with this judgement, and yet acts contrary to this. The closest that I can find to an argument for this claim in his discussion is that it provides a better account than Davidson’s of cases such as the following:

I decide, after considering all the facts known to me that bear on the issue, that it is best for me not to drink wine at dinner tonight. … [But the] wine being served looked rather tempting, and so in a moment of weakness, I drank it. (p. 227)

Davidson would say that in this example, though the agent made up his mind, hence wanted not to drink, his mind later changed—he may have retained the conditional all things considered judgement, but not his unconditional judgement. Searle’s response to this is, I think, correct: it seems eminently possible for the agent to have drunk the wine even though his unconditional judgement did not change. What is not so clear is that this is a case in which P1 fails. For though his unconditional judgement may have remained the same, why should we suppose that the balance of his wants also remained the same? After all, did not the temptation cause him to want to drink more than to refrain?1

The matter will be clearer if we pay attention to the different senses of wanting that might be in play. Perhaps there is a sense in which, along with the unconditional judgement, the agent still wanted more to refrain from drinking. But this, I suggest, would only be wanting in some respect. Though the judgment was unconditional, this does not imply that the agent's overall motivation was in favour of not drinking (assuming the falsity of P2). The temptation to drink caused him, for the moment, to be at least as motivated to

---

1See Michael (2008) for a more detailed discussion of how an agent can act contrary to an unconditional judgement.
drink as to refrain. That is, it brought about a balance of overall wants out of step with what he wanted in respect of his judgment. It is trivially true that an agent can act contrary to what he wants in some respect. But that is not the sense for which P1 is true.

There is another sense in which the agent could be said to have wanted more to refrain from drinking at the time of action: that of standing wants. In making up his mind, the agent formed a disposition to want not to drink. However, though the disposition was still present at the time of action, it was not being manifested. Again, it is trivially true that an agent can act contrary to his standing wants. But that is not the sense for which P1 is true.

To sum up, in criticising Davidson’s solution to the problem of weakness of the will, Searle is correct. But his inference that both P2 and P1 are false is not. Nothing he has said implies that P1 is false. Of course, as Davidson himself pointed out, there clearly are senses in which P1 is false, but to identify such does not show that there does not remain a pertinent sense in which it is true.

The Free Will Objection

To understand Searle’s position better, we need to put it into context. Much of Rationality in Action is spent arguing that there is a ‘gap’ in voluntary action. Searle defines this gap as ‘that feature of our experiences by which … we sense that the antecedent psychological causes of our action are not by themselves sufficient to determine the action’.  

His argument against Davidson is supposed to provide independent grounds for this claim— as he puts it, that P2 and P1 are false is ‘a kind of reductio ad absurdum’ of the view that ‘if the psychological antecedents of the act are all in order … then the act must necessarily follow.’ As we have seen, the argument fails to establish the falsity of P1, hence to provide independent grounds for the claim. This, however, would not unduly worry Searle as he thinks that the gap is in any case ‘an obvious fact of our conscious life’.  

However, I disagree about this also, as I argue in another paper. This is not the place to reproduce the argument in full, but a key part of it is the observation that Searle fails to distinguish between two things: the conscious experience that the psychological antecedents of an action are insufficient for it, and the absence of the conscious experience that the psychological antecedents of an action are sufficient for it. I agree that for most voluntary actions the latter holds, but I do not think that this implies the former. It is no more the case that the lack of experience of sufficiency amounts to an experience of insufficiency than it is that lack of evidence amounts to evidence of lack. What matters here is whether, if it is true that psychological causes are

---

1This is a formulation that Searle gives in his reply to Zaibert (2001, p. 84).
sufficient for action, we should expect the agent to experience them as such. But there is no reason to have such an expectation. Psychological causes do not operate externally to the agent but internally—they operate through the agent. So there should be no expectation, as Searle thinks there is, that if they are causally sufficient then the agent can sit back and watch their operation as if on a movie screen.¹

So both of Searle’s arguments for the existence of a gap fail and the principle P1 remains unharmed by his objections. Nevertheless there is something rescuable from his position. There is a simple alternative argument that can be brought against P1 which captures the spirit of Searle’s opposition to the principle. Regardless of whether or not it is based on our conscious experience, we have the belief that at any moment leading up to a voluntary action we are free to do otherwise. We believe that even if we are more motivated to x, as long as we are not under compulsion, we are nevertheless free to y instead. This is a very firm belief—arguably of a degree that trumps that of P1. Indeed, it is beliefs like this which constitute our notion of free will. Yet P1 appears to rule it out. So if we were to give up either belief, we should give up P1. This is the free will objection.

Does then P1 contradict free will? The answer depends on how one reads it. First note that P1 is in the form of a conditional whose consequent is itself a conditional. It does not say, as on occasion Searle seems to imply, that if one wants to x more than to y then it inevitably follows that one will x, but rather that it follows that one will x if one either x-s or y-s intentionally. Of course, this itself does not answer the free will objection, for it still seems to rule out the possibility of the agent y-ing. However, a sensible reading of P1 does answer the objection. This is that the latter conditional applies only while the agent wants to x more than to y. For it is surely the case that if the agent does not act while wanting to x more than to y, and then later changes to wanting to y more than to x, then the conditional no longer holds: there can be no necessity that one will x rather than y as long as one is currently more motivated to y than to x. On this reading, even while wanting to x more than to y, it remains perfectly possible for the agent to y as long as this does not happen while he wants to x more than to y.

However, even taking this into account, is it not still a violation of freedom to say that, whilst in the state of wanting to x more than to y, one cannot y? To see why it is not, consider the following analogy. Suppose I encounter an acquaintance of mine who is standing a few meters away and who, upon greeting me, holds out his hand to be shaken. Assuming that there is nothing preventing me from doing so, I am free to shake his hand. But I cannot do so while I am standing a few meters away. I need to get closer to him in order to shake his hand. That is, my shaking his hand requires a change in my physical state. Something similar holds for P1. Suppose I want to x more than to y. Assuming that there is nothing preventing me from doing so, I am free to y. But I cannot do so while I want to x more than to y. My y-ing requires a

change in my psychological state.

The analogy illustrates how one can be free to y even though one cannot y while wanting to x more. It suggests that our pre-theoretic notion of ‘being free’ to do something does not entail that the agent can do it while remaining completely unchanged. Remember, the claim that is supposed to threaten P1 is that whatever my current motivational state I am free to do otherwise. As stated, this is compatible with P1. What is not compatible is the claim that I am free to do otherwise while in the contrary motivational state. But we do not have the same intuitive grounds for endorsing the latter claim: it is the former claim that we take for granted.

This may not silence all doubts. For, some may ask, how does an agent’s psychological state change? Would he just need to wait until this happens? Can he bring about the change himself? How would he do this? It seems odd to say that he would need to wait for his psychological state to change, and at least a little off-beat to say that in the normal course of things he would need to self-consciously manipulate his own wants. So until we have answers to these questions the free will objection will remain attractive. To provide such answers I will first attempt to elucidate P1.

Elucidating P1

First we need to understand better the notion of ‘wanting’ at stake. Assuming P1 is true, what interpretation of ‘wanting more’ would make it so? The most obvious candidate is that of relative disposition. That wanting can be used in this sense has support from within ordinary English: ‘to be disposed to’ can be used as a synonym of ‘to want to’. Thus, for example, in the New International Version of 1 Corinthians 10: 27 we read, ‘If some unbeliever invites you to a meal and you want to go’, whereas in the King James’ Version we read, ‘If any of them that believe not bid you to a feast, and ye be disposed to go’ (my emphases). If wanting is interpreted as a disposition, then on a fairly straightforward reading of disposition, P1 turns out to be a conceptual truth. For if one were to expand on what it means to be (occurrently and overall) disposed more to x-ing than to y-ing, a natural answer would be, ‘faced by opportunity to do either, one would x if one were to do either.’

But there is more to wanting to x than just having a disposition to x: it is to have such a disposition in virtue of a positive representation of x-ing. If reference to an occurrent agential disposition towards action is to play a meaningful role in commonsense psychological explanation, it ought to serve as the basis of a deeper explanation. To understand the agent’s action one would need to know, not just that he is most disposed towards this action, but why he is so disposed. This requires reference to the grounds of the disposition. Though such are diverse, what they have in common is that they all involve representing the action in some positive light (hence the term pro attitude). We can say that the agent represents the action as desirable, obligatory, or in some other sense ‘to be undertaken’ (for convenience I will just say ‘represents as
desirable’). Thus, for example, one wants to go to the feast because one represents it as desirable to do so; one would not want to go to the feast if one did not represent it as desirable to do so.

Almost always an agent’s wanting to x is at least partly extrinsic, that is, is wanted not entirely for its own sake but for its conduciveness to something else that is wanted. One way to capture this in a general account is to say that the agent sees or construes x-ing in terms of some conjunction of features, F1…Fn, where a feature is a consequent state of affairs of x-ing (which may include x-ing itself) or a property of such, and that it is F1…Fn that is represented as desirable. Such construal need not be an all or nothing matter; it can occur in degrees, in other words, one can construe an action more in terms of one feature (or conjunction of features) than another. At any one time there will be many more features that the agent believes x-ing to have than he can construe it in terms of, so he can only construe it in terms of some of these. That subset of features that he does construe x-ing in terms of, and the degree to which he construes x-ing in terms of them, defines a perspective on x-ing. It is the current perspective that one is in that determines one’s occurrent disposition. As an agent’s perspective changes, so does his disposition. Crucially, an agent’s perspective can change as his thoughts change.

In order to see how this account can elucidate P1, let us return to the example of Bill that I gave earlier. Recall that Bill’s choice was puzzling because it is difficult to understand why he would want to take $1 rather than $100. There are, nevertheless, answers to this question. Suppose that initially when presented with the choice he wanted to take $100 more, just as one would expect, but then at the moment of choice he had the heroic thought, ‘I am a free agent and I will exercise that freedom by acting contrary to my desire,’ upon which he took the $1. Then the answer to the question ‘why would he want to take $1 more?’ is simply ‘because he saw it as an act of free will’. Though initially Bill had been seeing the options in terms of monetary gain, upon his thoughts of freedom his perspective changed: at the moment of choice he was seeing taking the $1 as an act of free will whilst taking the $100 as not and he preferred the former to the latter.1 So his thoughts changed his perspective, and his perspective determined the dispositional state out of which he acted.

The key point is that wanting, in the sense that makes P1 true, is a fluid state, one that varies as our representations of the options at hand vary. Thus it is not the case that, in order to act contrary to one’s current wants, one needs to self-consciously manipulate them. Rather, one is having thoughts all the time, and as this is taking place one’s wants may change. So if one has made up one’s mind to x but nevertheless, as Searle puts it, is finding y-ing tempting, then whatever thoughts are causing this temptation are changing one’s overall wants. We must not think of an agent as operating on his wants, in the relevant sense, but of those wants being a description of his occurrent state. As his thoughts change, so will his state

1One can also have scenarios in which he chooses from the position of a motivation tie.
Conclusion

In this paper I have offered both positive and defensive arguments in support of Davidson's principle. If I am correct in my positive arguments, then P1 is written into our commonsense psychological expectations, and the best interpretation of it as such is that wanting is a fluid dispositional state that changes as our perspective changes. If my defensive arguments are correct, then the belief that we are free to do otherwise does not entail that we can act contrary to what we want more without changing our state. Together I think these arguments constitute a strong case for accepting P1.

References

This essay serves to fulfill four basic aims; firstly, to respond indirectly to the protracted attack against metaphysics broadly construed, dating back to David Hume, culminating with the logical positivist movement. For the positivists, meaningful enquiries are drawn from one of two paths: either from empirical enquiries answerable to observations, or from ‘logical enquiries founded on relations of meaning’. It was construed that traditional metaphysical enquiries fall within neither of these two paths and consequently, its enquiries are ill-founded and metaphysics as such rendered a mere ‘self-sustaining game connecting with nothing in reality’ (Price, 1997, p.4).

Secondly, this essay builds on the impetus initiated from the recent publication of Metametaphysics (2009) edited by Chalmers et al. drawing on the insights and recommendations presented therein. My concern here is to demonstrate why a grounding in an ontology of binding relations is fundamental to revising ontological categories as a basis for our capacity to understand our world for both epistemology and metaphysics given that these two fields arguably converge on the quest to understanding ‘what it is - to know’. This aspect of the paper builds on the recommendations spelt out by Jonathon Schaffer (2009) that metaphysics is about ‘what grounds what’ by spelling out the structures of the world, about what is fundamental, and what derives from it. In this philosophers have much to offer the discourse of what there is.

Thirdly, indirectly, given the advance of quantum theory and physics in general, there is a need to review the manner in which the ontological categories of old require substantive revision as opposed complete rejection. Central to this process requires answering two fundamental questions both directly and indirectly respectively: 1) What is ontology? 2) Why is ontology of binding relations significant for the way in which we think about being, that is anything in the universe (real or idea-of)?

Fourthly, this paper attempts to dispel the ‘problem of identity and persistence’ and the associated ‘problem of composition’ when seen from the perspective of Space-Time-Event-Motion (STEM) concept, as the process of motion and change is built into the concept of composition through the constitution of this new perspective forming a principle of individuation.
What is Ontology?

Fred Sommers in *Types and Ontology* (1963) defined ‘ontology’ as the ‘science of categories’ operating as the systematic process of establishing what categories are, ‘how they are determined, and how they are related to one another’ (p. 351). George Englebretsen (1990) explains that a category is a group of things, not simply a class, but a special kind of class. A class consists of all things of which some given term is true. Simple enough, though establishing a robust and fixed ontology for class/es is nonetheless problematic. One of the major problems associated with the idea of classes is this notion ‘true of’ in relation to negation. As Sommers describes this problem pointing to contemporary logicians, that from the kinds of negation applicable in a natural language sentence, as amounting to the negation of the entire sentence. Sommers also made distinct what a term is ‘true of’ and what it ‘spans’ (1963, p. 329). Significantly, spanning is not defined in terms of truth, as a term and its negation are never true of the same things; however, they both (true and false) span the same things. So while classes are defined in terms of truth, less problematically, categories, Englebretsen (1990) elucidates are defined in terms of spanning (in turn defined in terms of sense). Further as Englebretsen (1990) interprets Sommers, a category regarding term ‘X’ will include the class of X things and some of the counter-class of X things (pp. 4-5). So the class of married things, for example, will include married members of that class. The counter-class will include anything of which married is not true. Importantly the category regarding married will include married members in addition to those members to whom the question of being married may sensibly be asked even though it may be false, for example, like my son Tristan being married (Tristan is not married). For while married is false of my son it spans only those individuals for whom the question could sensibly be raised as being married and not of other objects for example, Pluto or a Mulberry bush. A class of things may be subject to change over time and not be eternally fixed.

D.C. Williams, in *The Elements of Being* informs that First philosophy, in the traditional sense, ‘is analytic ontology, examining the traits necessary to whatever is, in this or any possible worlds’ (p. 57). Williams’ maintains that ‘metaphysics is the thoroughly empirical science’. He goes on to state that ‘… [e]very item of experience must be evidence for or against any hypothesis of speculative cosmology … and every experienced object must be an exemplar and test case for the categories of analytic ontology’ (p. 57).

In *Logic and Ontology* (2004) Thomas Hofweber reports that there are several conceptions of logic as there are several conceptions of ontology. Metaphysics, I contend, is a viable enterprise worthy of resistance against the current debate to reduce metaphysical concerns to scientific concerns since they deal with different types of questions. Though open dialogue among the respective practitioners should be encouraged. Kit Fine in *The Question of Ontology* (2009) states that the ‘commonly accepted view … is that ontological questions are quantificational questions’ (p. 158). Again however, reducing metaphysics to quantification or similarly other linguistic devices as, for
example, Quine has shown in challenging Carnap’s analytic-synthetic distinction, remains incomplete and so problematical because ‘linguistic rules are never absolute, and pragmatic restructuring is never entirely off the agenda’ (Price 1997, p. 8). Certainly one of the most constraining problems associated with quantification and predication that requires truth-bearing is ‘change’—for example, Aristotle (translated by E.M. Edghill) in Section 1 Part 5 of the *Categories*, taught that a statement can be both true and false. ‘The same statement ‘he is sitting is true, yet, when the person in question has risen, the same statement is false, [and] the same applies to opinions’ (p.6).

Change affects persistence. How then do we tackle problems such as the ‘problem of identity and persistence’, and the ‘problem of composition’? Karen Bennett explains: ‘some think composition never occurs — there are simples and that these simples have various properties, and stand in various interesting relations’ (2009, p. 44). Bertrand Russell (1948, Part IV, ch. 8) proposed a new version of property realism (rejected the traditional notion of substance) known as a ‘bundle theory’ since it regards ordinary objects as bundles of properties (universals) he called *compresence*. The immediate problem encountered by bundle theorists is how to provide an explanation for the ontology of individuals where issues like the problem of individuation and identity and persistence loom large (Swoyer, C. 2000). In relation then to Sommers’ analysis if nothing can sensibly be said that does not include the possibility of its negation or denial then it only seems plausible that there should be disputes about ‘what is’ and ‘what isn’t’. Life – Existence; Death its negation! Concepts like ‘attraction’ – ‘repulsion’ may mean ‘simples’ that bind (attraction) or not bind (repulsion); but this is not a sufficient analysis. One major stumbling block concerns the property realist’s insistence that universals must be immutable, however, this restriction may be overcome without affecting the substantive nature of a universal; particularly so if the universal is energy capable of instantiating multiple forms enabled by the conservation of energy principle. This important point is developed as we progress.

No serious philosopher would doubt that applicable questions of enquiry into the nature of the world emerge from the framework or methodology grounding the enquiry. Karen Bennett (2009) argues that many disputes among philosophers revolve around ‘difference-minimizing’, that is ‘…debates in which everyone takes the data to be largely the same. All the participants want somehow to preserve our ordinary judgments of persistence, of sameness and difference, of what there is and isn’t’ (p.72). She adds that not all metaphysical problems are difference-minimizing. Surprisingly from Bennett’s analysis her claim ‘… is not that work on the metaphysics of material objects is pointless, but rather that we have more or less done it already’ (2009, p. 73). I disagree; one need only look at the status of modern physics to appreciate the level of disagreement on various fronts (e.g. quantum mechanics) notwithstanding the plethora of philosophical disagreements.

I think one thing is obvious, that ontological categories, since Aristotle first enumerated their kind, arguably require a substantive role in philosophy/metaphysics to qualify above the ranking of a mere classification...
system. Classification I take to mean a systematic process of naming and grouping types of things in the universe, which inherently suffers incompleteness. Evidence of this view is still resonant as David Manley (2009) claims that ‘… metaphysics still has an important mission: to identify the best language in which to take inventory of the furniture of reality’ (p. 28). Prompting the question: Is the role then of metaphysics just the listing of apparent things in the universe? One may be inclined to deny metaphysics any role of listing all the elements in the universe insisting that it is the endeavour of predicate logic or quantification theory as long as the language employed is consistent. This presupposition seems however to be missing an important point about the purpose behind the discipline of metaphysics I dare say. Defending ontology built on classification alone, suffers incompleteness due to change and evolution. Any linguistic system suffers internal incompleteness subject to its own self-referential nature, requiring an internally structured proof domain to bolster its own constituent terms and grammar. It is no wonder then that there is no branch of philosophy that has suffered such protracted attacks against its viability than metaphysics and its sub-branch ontology if ‘taking inventory’ is all that is required of it. So how do we move forward?

We can move forward by examining how the simple idea behind Quine’s reform is captured by the claim ‘To be is to be the value of a bound variable’. In other words, ‘to be’ is ‘to be’ in some kind of binding relation, in this case, a relation of value as assigned in relation to a bound variable. To commit to ‘things that are F’ entails a quantificational sentence whereby establishing the truth requires ‘the existence of at least one object o that makes [Fx] true when o is assigned as [the] value of ‘x” (Soames, 2008 p.4). Hence an ontology of relations remains fundamental when examining how particulars and universals can exist, how to understand quantum entanglement, and in which its denial (relations) destroying the possibility of any kind of engagement even genuine connection of any sort would be rendered impossible. Indeed the very possibility of any sort of predication is realised, made cognizant necessarily, via ontology of relations. But of course there are several ways to think about the notion of relations.

On Relations

The concept of ‘relation’ has several modes of operation. I will present a version of ‘binding relations’ (connective role) drawn from Aristotle’s initial conception and sense of relation expressed in terms of ‘toward-ness’ (aditas). I will argue that this version is required in order to ground what would constitute the basis for at least some ontological categories. Currently in mathematics and epistemology there are several relation types such as ‘void relation’; ‘universal relation’; ‘asymmetric relation’ etc. these and more are based on the idea of one set. But in each case the catalytically combinatorial action required for binding is missing. Identity, for example, is commonly thought to be a relation involving reflexive, transitive and symmetrical properties. Any relation with
these three properties is called an ‘equivalence relation’. Nature exhibits a natural propensity for things to combine, e.g. atoms into molecules, cells into multicellular organisms. Electromagnetism is responsible for combining atoms into molecules that further combine to create the plethora of phenomena in the world. This is arguably a physical activity involving a compositional or binding relation process.

For our purposes an explanation of the Aristotelian concept of relation as ‘toward something’ is at the heart of this investigation. Our analysis in the first instance will be enriched by briefly exploring how the Empiricists made use of the concept of relation. David Hume is important in this context for his distinction between ‘Relations of ideas’ and ‘Matters of fact’ (Hume, D. 1910 Section 4 Part 1) and particularly for the significant role the ‘imagination’ serves in his philosophy. The imagination for Hume operates as the mechanism involved in bringing together impressions and breaking apart ideas to form new complex ones. The idea of relation also serves as an important platform in John Locke’s An Essay Concerning Human Understanding in which Book 2, Chapters XXV – XXVIII are entirely devoted to this concept.

*When the mind so considers one thing that it does, as it were, bring it to and set it by another and carry its view from one to the other, this is, as the words import, relation and respect; and denominations given to positive things, intimating that respect, and serving as marks to lead the thoughts beyond the subject itself denominated to something distinct from it, are what we call relatives; and the things so brought together related” (Locke p.168).*

The idea of relation is indeed pivotal in Locke’s philosophy and it appears the greater sense of the connective role it served escaped his own scrutiny. Locke argued that the mind is a *tabula rasa* filled by experience in the form of sensation and reflection which provide the basic materials, simple ideas, from which most of our more complex knowledge is constructed. It is in the second action of the mind that ideas of relation emerge when the mind brings together two ideas whether simple or complex so as to take ‘a view of them at once’ (Locke, p. 168) without uniting them (Uzgalis, W. 2007, p. 11). What one gleans from this analysis is a connective or binding activity of the mind dependent on ideas of relation such to provide a new idea. Indeed a new way of seeing the world and of understanding. Remarkably not much has been said about this important idea of binding relation.

Over sustained periods medieval philosophers argued about the nature and ontological status of relations that provides much insight for this research. An excellent treatment and thorough analysis of this area of medieval philosophical debate was recently conducted by Francisco Suárez and John P. Doyle (2006). Medieval philosophers, in their analysis of Aristotle’s work explored the division of relation into real and rational.

*Relation is usually divided into that which is real or that which is*
only rational, which some interpret such that they teach that the categorical genus, ‘toward something’ (Ad aliquid) includes both relations and that this division is therefore univocal, indeed that it is a division of a genus into its species.

In order to establish what can be regarded as real in the world distinct from thought real, an examination of how the very concept of relation as ‘toward something’, as determined by Aristotle became the subject of further investigation.

... those things are said to be relative whose whole being is such as “to have itself toward another.” But this is entirely and most properly fitting for relations of reason. For although the being of these is more imperfect and diminished than the being of real relatives, nevertheless, their being, such as it is, consists entirely in a disposition (habitudo) toward something else, just as properly as does the being of a real relation. Thus it turns out that, just like a real relation, a relation of reason can be neither known nor defined without a disposition to something else. ... Nevertheless, we have to say, that only real relations make up the category, “toward something” (Suárez, F. Doyle, J.P 2006 pp. 79-81).

The idea of relation ‘toward something’ is what makes this type of relation useful for the type required for compositional binding (connective-relation). So significant is the idea of relation ‘toward something’ as demonstrated by Suarez and Doyle in their examination of this medieval investigation as to warrant, the authors interpret, a separate ontological category for the relation ‘toward something’:

[When a relative is said in general to be that whose being is to be toward another; it is necessary to understand in that definition, “to another that is correlative.” Therefore, since the same thing cannot be related to itself by a real relation (about which we are now treating), it will be necessary that there is not one supreme genus, but rather two, which respect each other adequately and are in reality (in re ipsa) distinct (Suárez, F. Doyle, J.P 2006 p. 245).

Perhaps what is most significant about this argument is the sense it provides for a systems analysis of how distinct objects, organisms so related, affectively, not only shapes habitats and ecosystems, but specifically living organisms whose various compositional systems are so integrated. Steven Rose (1997) reminds us that when we act with our environment we change it and are changed by it in this reciprocal relation. Interestingly, there is a direct correlation here to the fundamental nature of phenomenological enquiry in that consciousness is always directed toward something.
Conceptualising Grounding Categories

Jonathan Schaffer (2009) argues that the debate between Quine and Carnap pivots around their mistaken assumption that metaphysics is about existence questions such as whether numbers exist, etc. but as Schaffer points out with Aristotle the discussion is about substances (fundamental units of being). ‘For Aristotle metaphysics is about what grounds what’ and substance is its first part which he developed in the *Categories* (2009 p. 351-52). Ontological priority is the main criterion for selecting the primary substances and an entity is ontologically primary if other things depend for its existence on it (p. 351). Schaffer reminds us that metaphysics is the discipline that studies substances and their modes and kinds, by studying the fundamental entities and what depend on them (2009 p. 353). Schaffer acknowledges that existence questions do play a role for his type of neo-Aristotelian philosophy. He suggests that what exist are the grounds, grounding relations and the grounded entities (2009 p. 353). Categories are significant because they serve as ‘places in the dependence ordering; substances, for instance, serves as both root node and focal category’ (2009, p 356). The target of metaphysical inquiry, then, Schaffer argues, is an ordered hierarchy generated from, (i) a list of substances F, plus, (ii) a list of grounding relations G, (2009 p. 363). Grounding criteria he announces should be irreflexive and asymmetric (2009 p. 374). The type of ontology I am arguing for structurally meets Schaffer’s requirements.

In the following section an argument is advanced on what reasonably should be regarded as one of the primary substances and how to think about grounding relation i.e. binding relation. The subject we must first deal with is energy construed as a primary substance and not simply as a contingent property. This is a hot topic among physicists and philosophers of science and not an issue that can be completely resolved here. However we may add observations that indeed shed new light on the subject.

The Grounding of a Primary Substance

Could ‘energy’ be a grounding primary substance? Should energy be thought of as just a property or something more substance-like? Energy is commonly understood as ‘the ability to do work’ which is simple enough to understand. But energy strictly defined in such a manner I contend misconstrues the concept affectively constraining the phenomenon to a property like term. For example, substances putatively understood always have heat energy expressed by the activity of vibrating atoms. Vibration in this sense should properly be construed as a property of atoms. Another example is water. Is water a substance? No, because water is a property of a collection of H₂O molecules. Take apart the molecules and atoms, delve inside the sub-atomic particles and you will not find water anywhere because water is just a property of something else. Physicists indeed hold that any property which is a ‘conserved’ property is also a substance-like entity. As such, energy seems to be
much more fundamental, indeed substance-like.

In earlier work (2002) I grappled with the concept of energy to develop an understanding from a different angle. The concept of energy is an elusive term that required a richer metaphor in order to penetrate the conceptual extremes and multiplicity of frameworks it features in the physical sciences. I developed the conjoined concept Space-Time-Event-Motion (STEM) (a concept for concentrated energy underpinning human form) to metaphorically extend beyond the particular but also to be employed as an intransitive verb, (not governing an object). This idea can now be thought of as a spanning notion for individual entities.

STEM refers to a biophysical process (animation) that incorporates the change and motion of an event whose measured duration can be coherently defined given the context of acquisitional requirement i.e. definitional, explanatory or experimental work. That is, processes observed in causal terms within a region (I refer to as a containment-field, an embodying, and localising principle) would exact differing levels/degrees of energy and hence, embrace the core notion of STEM. Why? The answer is that time is associated with the mass of an object, which is the measure of inertia whose influence is directly responsible for physically affecting the curvature of space-time as defined in relativity theory. Significantly, the degree of curvature of spacetime is thus proportionally related to the inherent energy (mass) of an object in question. Matter, according to relativity theory is reducible to energy (i.e. E=MC^2). Arguably, whatever energy is, it seems to be the most pervasive stuff in and of the Universe. Indeed time, then, is a constitutional part of the very fabric of the Universe. Spacetime in-forms mass by way of how to move and mass in-forms spacetime by way of how to curve. The fabric of the Universe, at least in an empirical sense, is fundamentally physical (The term ‘physical’ I take to mean anything that can be acted upon in a causal manner.). If time is physical, it is reducible to energy according to relativity theory. If time is not physical, it cannot have any effect on the physical objects of the Universe! The Universe is made up of physical things (if not completely at least in part). Ergo, time is physical.

The question we must now answer is: what is required to construct an individuation principle? Particularly poignant when confronted with ‘the problem of identity and persistence’ and the ‘problem of composition’ since most thinkers engaging in these issues have presumed an entity exists in space and is subject to change construed in temporal terms. Consequently, I contend the presupposition is ill-founded. Address these same problems from the STEM perspective and immediately they are not so intractable, indeed, I dare say, they dissolve. My point is that phenomena, i.e. humans and other individuals are made up of STEM as part, if not the whole, of their constitution. Spatiotemporal change is constitutional because all entities as such are in motion. The simple analysis is to make this observation. An individual’s life is an existent event (duration of one’s life) in physical terms. Motion is integral to life. So from the above construction, a living, existing, individual is a STEM causal agent.
This STEM concept is formed from what I am arguing is one type of binding-relation. This STEM concept is relevant when referring to organisms exhibiting self-organising capacities, like human beings, such that the expression of activity or the process of an entity’s life is a relation ‘toward-something’ as being in engagement with its external and internal environments. For Aristotle this is akin to the notion of ‘potency’ dealt with in Book V Chapter 12. Aristotle says that ‘[p]otency means a source of motion or of change, either in something else or in something as something else’ (Aristotle, p. 92). Yet from this analysis thus far, as many before me have proclaimed, that the ontological categories Aristotle pronounced need revision to account for the way in which the world exhibits itself in the 21st century. How are we to understand the notion of substance if it is not subject to empirical investigation? Let us explore this from another angle that takes us closer to how this all relates to energy.

Time and measurement are handmaidens and there is a direct relation between these concepts and energy. The most accurate measuring procedure is the atomic clock based on the regular vibrations associated with atoms and the standard form of atomic clock is based on caesium atoms. That is ‘…the spectrum of caesium includes a feature corresponding to radiation with a precise frequency – 9,192,631,770 cycles per second’ (Gribbin, 1998:29). One second is defined as that amount of oscillations of radiation. The caesium clock is one part in 10\(^{13}\) (one in 10,000 billion), or one second in 316,000 years accurate (Gribbin, 1998:29). In physics, ‘Planck’s constant’ is described as the constant proportionality between the energy emitted or absorbed by an atom and the frequency of emitted or absorbed light as an electromagnetic wave (Jibu & Yasue, 1995). Planck’s constant suggests that energy and frequency and consequently ‘time’, as far as measurement is concerned, are all interrelated concepts. John Maddox (1998) concludes that the greater the frequency, the greater the energy of the quanta. This equates then to what I have termed Signature-Energy-Frequency (SEF). Further, Einstein’s theory of relativity demonstrates that time is a relative concept and that travelling at near speeds of light produces ‘time-dilation’, a slowing down of time. Motion is an important conjunct in this observation. Matter and its properties such as ‘mass’ and ‘charge’, according to the theory of relativity are all reducible to energy. Therefore, energy is indeed the fundamental substance of the Universe definable in terms of frequency. Thus, the adjoining concept Signature-Energy-Frequency (SEF) operates on both the quantum and classical levels that form STEM containment-fields, metaphorically subsuming the concept of the principle of individuation (see Naimo, J. 2002 & 2009).

**Concluding Remarks**

The main thrust of this paper was to demonstrate the manner in which the problem of identity and persistence or composition seen from a differing framework can provide much needed light and perspective. I introduced a new
An Anthology of Philosophical Studies

type of relation namely a binding-relation as a means to think of STEM as the constitutional fabric of individuation. An argument was advanced suggesting that energy should be considered one of the primary substances that form the matrix of establishing relevant ontological categories.

References

Chalmers, DJ; Manley, D; and Wasserman, R. (Eds.) (2009) *Metametaphysics: new essays on the foundations of ontology*, New York, Oxford University Press,
Sommers, F. (1963) ‘Types and Ontology’ in *The Philosophical Review* 72 (pp. 327-363)


Distinguishing mathematics is its method, deduction being a defining characteristic. Inheritance is at least from “Euclid [who] sought to organize geometry in a systematic deductive form.”\textsuperscript{1} Minimally, this was done with “the aim . . . of proving beyond all reasonable doubt that certain laws of geometry hold.”\textsuperscript{2} To achieve this, mathematics proceeds by “inference in which the conclusion about particulars follows necessarily from general or universal premises.”\textsuperscript{3}

Since, “Mathematics is a deductive science: starting from certain premisses, it arrives, by a strict process of deduction, at the various theorems which constitute it.”\textsuperscript{4} Descending from Classical Greece, Euclidean geometry dominated Western mathematics because mapping the observable universe necessarily. Kant’s first antinomy\textsuperscript{5} revealing it contradictory, George Boole supplanted geometry with algebra.

Kant did understand Euclidean geometry as real, a synthetic a priori category. Identified is an axiomatic system whose theorems are deducible from its axioms, and whose axioms are inducible from its theorems. Self-determining like the Platonic dialectic, it is real, not nominal.

Assuming Kant’s first antinomy, however, Euclidean geometry contains an inconsistency. If something is infinite, it is unlimited. If something is unlimited, it is indistinguishable. If something is indistinguishable, it has no being. If something has being, it is distinguishable. If something is distinguishable, it is limited. If something is limited, it is finite. Therefore, the infinite has no being. Assuming, “A piece of straight line may be extended indefinitely,” when indefinite extension is infinite, Euclid’s Fifth Postulate incorporates a contradiction, a being that has no being.\textsuperscript{6}


\textsuperscript{2}Barker 24-25.


Excluding lines which “may be extended indefinitely,” Nikolai Lobachevsky’s Parallel Postulate can be formulated, rendering Euclidean geometry nominal, not real.¹ Even in his “diagonal proof,” Georg Cantor assumes a limited “set [Inbegriff] M of elements” limiting its power set.² Different consistent geometries being self-contained, they need not be mutually consistent. Euclidean geometry can draw a line, but not a curve. Reimannian geometry can draw a curve, but not a line. Although intuitive, then, contrary to Kant’s presumption, Euclidean geometry is not necessary.

Kantian categories compose Boolean algebra, avoiding ambiguity by the uniform content of a mathematical field.³ Sacrificed, however, is the multiform content of complex relations. Supposing,

It is possible to assign relations among them [the classes composing mathematical content], whether as respect the repetition of a given operation or the succession of different ones, or some other particular, which are never violated, requires the classes composing mathematical content to share common constituents.⁴ Doing so, mathematical classes are ambiguous, indeterminate as to whether joined by common constituency or not.

Wedded to the deductive method whose necessity requires consistent categorical sets, confronting modern mathematics is the challenge of generating such sets without contradiction. Cognizant of the ambiguous character of the Kantian categories, Bertrand Russell abandons a wholly synthetic a priori mathematics. Manifest in the form of tautology, Boole’s classes cannot constitute premises in “a strict process of deduction” because they compose self-contained sets. Why so, “Russell . . . . formulated in the slogan, ‘If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total.’”⁵ Premisses in “a strict process of deduction” cannot be self-contained, rationality being derivative.

Presumably adhering to, “that feeling for reality which ought to be preserved even in the most abstract studies,”⁶ when concluding a collection has

---

⁵Barker 85.
no total, Russell indicates a “collection” is not a “total.” Implied is a “collection” is real and a “total” is unreal.¹ Not “definable in terms of” itself, a total is “definable in terms of” a tautology.

Specifically.

The importance of ‘tautology’ for a definition of mathematics was pointed out to me by my former pupil Ludwig Wittgenstein, who was working on the problem. I do not know whether he has solved it, or even whether he is alive or dead.²

As for himself, Russell concedes, “For the moment, I do not know how to define ‘tautology.’” Resolution is provided concluding totals only need be imaginary, not observational, so they are “real but have their reality only within the mind.”³

Abandoning possibility of a self-determining tautological axiom, while avoiding the indeterminacy of the infinite proof of Russell’s theory of types, Ian Stewart supplants the logical with the psychological. Thus, “At any stage in mathematics, one’s definition of ‘logically rigorous’ tends to boil down to ‘it convinces me’; though of course a professional logician takes a lot of convincing!”⁴ Why a “professional logician” is so recalcitrant, and recalcitrance a virtue, is left unsaid, convincing being accidental.

Sacrificed by psychologism is, “In mathematical subjects a fully deductive treatment is aimed at. The ideas are supposed to be so clearly defined that one can develop them by a purely logical argument.”⁵ Mistakenly, “In fact, to a large extent it is true that mathematical progress has been made only with problems which are in one sense or another linear.”⁶ Impossible is a proof whence,

If \( p \) is some statement in the system, a proof of \( p \) consists of a sequence of statements, each of which is either an axiom or a logical consequence of certain \textit{preceding} statements in the list, such that the last statement in the list is \( p \).⁷

\[ \text{Axiomatics} \]

Experientially indistinguishable from the psychological, distinguishing the

---

³Barker 31.
⁴Stewart 9.
⁶Sawyer 112.
⁷Stewart 117.
logical is the universality of the experience among human beings. Distinguishing the psychological is the particularity of the experience among human beings. Indicative is to be deviant, vary from a statistical norm, is *pathological*, diseased. To not be deviant, not vary from a statistical norm, is *normal*, healthy.

Unsurprisingly disease is designated “morbid,” generally repulsive to humans, and health is designated “wholesome,” generally attractive to humans. Presumed is humans are repulsed by other humans whose normal nature is incomplete (*unwholesome*), and attracted to other humans whose normal nature is complete (*wholesome*). Expectations for other species vary accordingly. Thus, failure of a female praying mantis to behead her male partner during coitus is pathological, whereas success of a female human being to behead her male partner during coitus is pathological, although once when using this illustration in a class, a female student retorted, “Go girl!”

Linear argument initiates with David Hilbert’s “distinction . . . between a subject matter under study and discourse about the subject matter,”¹ and continues with Ernst Zermelo’s “‘axiom of choice’” stipulating,

for any set whose members are sets that are non-empty and mutually exclusive, there exists at least one set having exactly one element in common with each of the sets belonging to the original set.²

Mathematically identity of “one element in common with” “a subject matter under study and discourse about the subject matter” proceeds whereby

In a similar fashion, a unique number, the product of as many primes as there are signs (each prime being raised to a power equal to the Gödel number of the corresponding sign), can be assigned to every finite sequence of elementary signs and, in particular, to every formula.³

Mathematical development employs, “AN ALGEBRA OF LOGIC”⁴ wherewith,

If any collection of these symbols is written, we judge the result to be true, false, or meaningless by using the vocabulary given . . . . We replace each symbol by the word or words it stands for, and pass judgement of the resulting sentence.⁵

Challenge now appears when argued is Kurt Gödel proved two things:

---

²Barker 77.
³Nagel and Newman 72.
⁴Sawyer 90.
⁵Sawyer 91.
On the Nominal Constitution of Mathematics

1. If axiomatic set theory is consistent, there exist theorems which can neither be proved nor disproved.
2. There is no constructive procedure which will prove axiomatic set theory to be consistent. . . .

Later developments have shown the wreck to be greater than even Gödel imagined. Any axiomatic system sufficiently extensive to allow the formulation of arithmetic will suffer the same defects. It is not any particular axiomatization that is at fault, but arithmetic itself.\(^1\)

For any nonempty set, it always is possible to imagine a member constituent of another set. Ernst Zermelo’s “‘axiom of choice’” manifests this.\(^2\) Such an element constitutes ambiguity, concurrently constituent of two or more “non-empty and mutually exclusive” sets. Constituent of “mutually exclusive” sets, it is a limit. Not constituent of “mutually exclusive” sets, it is not a limit.

“Mutually exclusive” sets can have “one element in common,” without constituting the common content of “one set” only if identifying the “one element in common” is an intrinsic and not a relational property.\(^3\) Identifying relational properties, elements of one set within which it is constituent are relatable by it to elements of every other set within which it is constituent, constituting one set within which all elements of the “non-empty and [mistakenly identified] mutually exclusive” sets are constituent. Indistinguishable as defined by intrinsic properties alone or relational properties also, whether an element occurs “in common with each of the sets belonging to the original set” is ambiguous.

Resolution of this ambiguity requires appeal to a nominal criterion constitutive of revisionary metaphysics. Now philosophy encompasses,

*The Criterion Argument*. We cannot distinguish truth from falsity without a criterion or rule. Of course, that criterion has to be true. Either we have no reason to think that it is true or we must introduce a second criterion that guarantees *its* truth. [W]e have entered on an *infinite regress*.\(^4\)

Foundation of axiomatic theory is laid, whereby how we speak is a function of “a criterion or rule.” Thus, “What we cannot speak about we must pass over in silence,”\(^5\) only relative to another axiomatic system, and this not absolutely considering David Hilbert’s “distinction . . . between a subject matter under study and discourse about the subject matter.”

\(^1\)Stewart, 292.
\(^2\)Barker, 77.
\(^5\)Ludwig Wittgenstein, *Tractatus Logico-Philosophicus* (Atlantic Highlands, New Jersey: The Humanities Press Inc., 1974), proposition 7, 74. This gives meaning to Socrates’ conclusion, “One thing only I know, and that is that I know nothing.”
Induction

Relevantly, “In the . . . so-called axiomatic method, an arbitrary, usually finite, set $X$ of sentences—an axiom system—is given, and the set $Cn(X)$, i.e. the smallest deductive system over $X$, is formed.” 1 Constituted is, “An ordered set $M$ [which] is said to be well-ordered, if $M$ itself, as well as every nonempty subset of $M$, has a first element under the order prescribed for its elements by $M$.” 2 Difficulty occurs in identifying “a first element.” Discourse about “a subject matter under study” necessarily occurring in the domain of another axiomatic system assuming “the sign for a function already contains the prototype of its argument, and it cannot contain itself,” 3 this identity occurs by a one-for-one substitution of symbols in one language for another language. Doing so,

Gödel first showed that it is possible to assign a unique number to each elementary sign, each formula (or sequence of signs), and each proof (or finite sequence of formulas). This number, which serves as a distinctive tag or label, is called the ‘Gödel number’ of the sign, formula, or proof. 4

Any “criterion or rule” being composed of elements, any element is deducible from its conjunction with another element, and its conjunction with another element is inducible with another element. Therefore any “criterion or rule” is deducible from conjunction of its elements, and is inducible with conjunction of its elements. Identity of a rule from its elements always being possible, and language being a rule system,

no society has ever known its language to be anything other than something inherited from previous generations, which it has no choice but to accept. That is why the question of the origins of language does not have the importance generally attributed to it. 5

False is, “Ethics so far as it springs from the desire to say something about the ultimate meaning of life, the absolute good, the absolute valuable . . . does not add to our knowledge in any sense,” 6 speculative revisionary metaphysics

---

3Wittgenstein, Tractatus Logico-Philosophicus, nt. 3.333, 17.
4Nagel and Newman 69.
5Ferdinand de Saussure, Course in General Linguistics (La Salle, Illinois: Open Court, 1986), 72.
being requisite for *identity* of any axiomatic system.\(^1\)

Logic is understandable as identity, a proof determining identity, whether deductively identifying particular as constituent of general, or inductively identifying general as constituted by particular. Imagining the set of all deduction, it is identifiable only inductively. Imagining the set of all induction, it is identifiable only deductively. Induction identifying deduction and deduction identifying induction, they are mutually identifying.

Thus, “a fully deductive treatment” does not impart necessity to the conclusion of mathematical argument. Necessity is imparted only when deductive argument is inductively identified. Deduction is understandable as a necessary condition of mathematics insofar as it is understood to not impart necessity to mathematical conclusions. Mathematical intuitionists are wholly cognizant of this.

They are understandable to presume mathematics axiomatic, and constitution of an axiomatic system to be mathematics. Assuming mathematical conceptualism, objects of mathematical knowledge “have their reality only within the mind.”\(^2\) With this, mathematical intuitionism concludes,

> we must possess a constructive proof of any mathematical statement about numbers before we are entitled to say that we know the statement is true. . . . we must know how to construct, or compute, such a number, using only a finite number of steps.\(^3\)

Constitution of an axiomatic system being inductive, then induction is constituent of mathematics. Induction is iterative. Iteration resolving ambiguity, and ambiguity being contradiction, then without iteration mathematics contains “concealed contradictions.” Enumerating a finite number of axiomatic elements is assumed eliminating contradiction. But axiomatic elements do not appear the same in every instance of their occurrence, symbols written differently, “tables, chairs, beermugs” appearing differently.\(^4\) Thus, occurrence of axiomatic elements is ambiguous, reintroducing the “concealed

---

\(^1\)Earle, 99.

\(^2\)Barker, 31.

\(^3\)Barker, 75.


> One result of this orientation is that anything at all can play the role of the undefined primitives of points, lines, planes, and so on, so long as the axioms are satisfied. Otto Blumenthal reports that, in a discussion in a Berlin train station in 1891, Hilbert said that in a proper axiomatization of geometry “one must always be able to say, instead of ‘points, straight lines, and planes’, ‘tables, chairs, and beer mugs.’” (see Hilbert 1935: 388-429; the story is related on p. 403)

contradictions” of which mathematical intuitionists warn.\(^1\)

Support for “concealed contradictions” can be provided arguing, since just as element can be conjoined with element to generate “an infinite regress,” the regress itself can be conjoined with its contradictory, \(A \land \neg A\), denying every constituent of the regress tit-for-tat, \(a_1 \land \neg a_1, a_2 \land \neg a_2, \ldots\), rendering the regress vicious because \(a_n \land \neg a_n\) satisfies the condition of membership in \(A\), which is being \(a\). Eliminating such a possibility requires a new rule, but this rule can be conjoined with its contradictory, \(\{A\} \land \{\neg A\}\), etc., rendering any “regress involved in predication” vicious.

Challenging this, Charles Saunders Peirce adjudges,

> All phenomena therefore are inferences or interpretations, and so are to some degree conceptual. Combining this fact with the apparent regress involved in predication, Peirce drew the conclusion that all cognitions are the results of inference, and so are logically determined by prior cognitions of the same object. The fact that this involves an infinite regress Peirce accepted once he was convinced that the regress was not vicious.\(^2\)

Such because only exclusive disjunction \(a \land \neg a\) is contradictory, inclusive disjunction \(a \land \neg a\) is consistent. Inclusive disjunction constitutes the conditional \(a \rightarrow \neg (\neg a) \land \neg a \rightarrow \neg (a)\), or \(a\) unless not \(a\), and not \(a\) unless \(a\). This is equivalent to the disjunction \(a \land \neg a\), the truth tables for both propositions being the same, the propositions being indistinguishable in their verification conditions. Conjoined disjunctives by this means are understandable as mutual qualification, a casuistic clarification of the circumstance in which each occurs. Now conjunction of \(a\) and \(\neg a\) is consistent, contradiction endlessly supplanting qualification and qualification endlessly supplanting contradiction, interminably devolving toward the unique identity of casuistry. Nominal, not probable representation of real, casuistry generates an algebraic unknown number, arbitrarily endlessly resolving ambiguity determining a set limit.\(^3\)

Concerned with E. J. Lowe’s “relational properties,” Wittgenstein seeks to preserve distinction between an element in a domain and an element in a converse domain by declaring equivalence.\(^4\) Still, Wittgenstein points the way to solving Frege’s Puzzle, while acknowledging like Russell “I can’t myself

---

1 Nagel and Newman, 22-3.
3 It is portrayal of casuistry as “1 : [an objective] resolving of specific cases of conscience, duty, or conduct through interpretation of ethical principles or religious doctrine,” which renders it, “2 : specious argument : RATIONALIZATION.” ["casuistry," in Merriam Webster’s Collegiate Dictionary, 178.]
4 Lowe, 105.

280
say quite clearly yet what tautologies really are.”¹ Wittingly or unwittingly, Russell completes the solution when recognizing, “Between two perceived perspectives which are similar, we can imagine a whole series of other perspectives, some at least unperceived.”²

Wittgenstein and Russell’s befuddlement as to the nature of tautology is resolved now by explaining identity a transmutative topological identity of all disjunctives of every power of the archetype and an equivalent power of the autotype. Elements in different domains being different, at least in relationship to domain elements, how are they identifiable as same, establishing “a one-one relation?” This is Wittgenstein and Russell’s difficulty with tautology. Identity is possible only if there is an unbroken identity of all disjunctives of elements in different domains. Requisite is an inductive iteration of an element in a domain terminating when fused into an indistinguishable whole with an element in a converse domain.

Constituted is a topological transmutative mapping of elements into integrated domains. Elements in different domains are proven related by tracing a continuum between them, incorporating both into a common domain. Proof is tracing such an undivided sequence, mathematically constituting identifying a dense set. It is material when physical, a continuity of matter between limits. It is mental when phenomenal or conceptual.

Engendered is an algebraic unknown number. This is formally indistinguishable from an irrational number, except progressing to the left rather than regressing to the right of a decimal point. Although substantively rational, an unknown number is indeterminate like an irrational number. Identified are all resolutions of Russell’s “Between two perceived perspectives which are similar, we can imagine a whole series of other perspectives.” Endless in extent, generated is an unascertainable decimal number. Because an arbitrary employment of exclusive disjunction endlessly, an unknown number need not be generated to know it exists.

Generation as an unbroken continuity eliminates concealed contradictions within any discontinuity, such contradictions presenting exclusive disjunctives separating rather than relating elements. Understanding inconsistency at a lower level of scale with a higher as qualifying the higher eliminates concealed contradictions within an unknown number, although the sequence is unknown. Provided is a casuistic determination of identity by an endless process of clarification.

Accepting “modern logic was designed with the language of mathematics largely in mind,” when “In mathematical subjects a fully deductive treatment is aimed at,” like George Boole, Kurt Gödel concludes mathematical “Classes and concepts [are] real objects.” Difficulty occurs considering the observational manifestation of any mathematical class or concept is indistinguishable from the observational manifestation of its contradictory. Thus, mathematical classes and concepts are nominal, not “real.”

Conceding mathematics is an axiomatic system, beginning from the locus of its axioms, mathematics is derivatively a necessary deductive system. This progression is equivalent to a numeric sequence to the right of a decimal. However, beginning from the locus of its axioms, mathematics is constitutively an unnecessary inductive system. This progression is equivalent to a numeric sequence to the left of a decimal. Inductively generated is endless arbitrary qualification of mathematical identity exhibited in continual proof of its axioms.

Speculatively Wittgenstein acknowledges deductive identity with, “Whereof one cannot speak, thereof one must be silent.” Inductive identity is acknowledged with, “‘Now I can go on!’” when “cannot [a] hearer still interpret . . . differently?” Introducing linguistic pragmatics, to “go on” succumbs to “The dark side” of “metaphor,” generating Wittgenstein’s “family resemblances.” Implemented is Franz Brentano’s concept of intentionality with which, “we first mark out a domain of relations-in-intensions by defining a construction process for the corresponding sentence matrices.”

Gottlob Frege translates intentionality into propositional operators, which are themselves translatable into inclusive or exclusive resolutions of disjunctives of limiting elements. Resolved exclusively, a disjoint is not Zermelo’s “element in common with each of the sets.” Resolved inclusively, a disjoint is Zermelo’s “element in common with each of the sets.” Being related to the other elements in each of the presumed sets, a disjoint’s inclusion in each

---

3. Wittgenstein, Tractatus Logico-Philosophicus, nt. 7, 108. See also nts. 5.6 and 5.61, 88-89.
7. “. . . psychologists in earlier times have already pointed out that there is a special affinity and analogy which exists among all mental phenomena, and which physical phenomena do not share.” Franz Brentano, Psychology from an Empirical Standpoint, trans. Antos C. Rancurello, D. B. Terrell and Linda L. McAlister (London: Routledge, 1997), 88.
of these relates all of the elements in them, constituting a common content to the sets, fusing them into a single set.

Pursuant to set integration,

we must possess a constructive proof of any mathematical statement about numbers before we are entitled to say that we know the statement is true. . . . we must know how to construct, or compute, such a number, using only a finite number of steps.

Hereby L. E. J. Brouwer implements Frege’s propositional operators within mathematical intuitionism.¹ “Frege’s Puzzle: how, if true, can ‘A = B’ differ in cognitive significance from ‘A = A’?” is resolved by the topological transmutation of A into B.² Considered as inconstant aspects, A = B. Considered as constant instance, A = A.

Relevant to defining this continuum are the concepts of “closeness” and “neighborhood.” Every member of a set of elements is a member of a set of subsets of that element in relation to every other member of the initial set of elements, identifying a topological “system of neighborhoods at x.”³ Hence, “In a certain sense, a neighborhood of a point x is a set of points which lie ‘close’ to the point.”⁴ “‘Close’” is qualified to indicate its inherent ambiguity. Intermediate of the limits of any sequence, geometrical or numerical, are an infinity of elements. Space is infinitely divisible, and number is infinitely divisible assuming,

By considering the convergence of the sequence \([0.999 \ldots = 1]\), we can show that the magnitude of . . . difference must be smaller than any positive quantity, and it can be shown . . . the only real number with this property is 0. Since the difference is 0 it follows that the numbers 1 and 0.999 . . . are identical.⁵

Thus, no matter how “close” any two elements are in a neighborhood, they

---

⁴Baum, 20.
⁵In mathematics, the recurring decimal 0.999…, which is also written as $0.\bar{9}$, or 0.(9), denotes a real number equal to 1. In other words, the notations “0.999…” and “1” represent the same real number.

always can be “closer,” until fusing into an indiscernible whole, rendering “closeness” meaningless.

“Closeness” is necessarily nominal, therefore, a function of “a constructive proof.” Elements can be equally close as well, if defined so. Indeed,

on Brouwer’s view, there is no determinant of mathematical truth outside the activity of thinking, a proposition only becomes true when the subject has experienced its truth (by having carried out an appropriate mental construction); similarly, a proposition only becomes false when the subject has experienced its falsehood (by realizing that an appropriate mental construction is not possible).1

Topologically, then, elements of a sequence are constrained by mutual relation—“closeness” and “neighborhood.” More than one point being “close” to a point in a “neighborhood,” “in an arbitrary topological space, limits of sequences need not be unique.”2 A topological sequence can achieve the same limit in different sequences of “the elements of the set we are studying,” whereby “two topologies may be alike.”3 Thus topology models the logic by which a mathematical axiom system is inductively generated.

Conclusion

A “language . . . inherited from previous generations” is equivalent to beginning with points distributed amongst cells of a topological neighborhood grid. Of this “one cannot speak,” but “can go on.” Tracing a path in the grid from cell to cell maps metaset to set in a transforming continuum resolving Frege’s Puzzle. More than one cell being “close” to another cell, this path is nominal,4 in which sense “logically rigorous” boils down to ‘it convinces me.’” Converging upon a common axiom system, different paths generate the same limit, illustrating the nominal inductive character of “The axiomatic method [which] consists in accepting without proof certain propositions as axioms or postulates.”5

---

3Baum, 21.
4Acknowledging “in an arbitrary topological space, limits of sequences need not be unique,” topologists proceed to exclude this possibility by introducing “Hausdorff (or T2) space.” [John D. Baum, Elements of Point Set Topology (New York: Dover publications, Inc., 1964), 40-41.]
5Nagel and Newman 4.
In this paper I propose a theory of folk psychology that bridges the metaphysical and semantic differences between the eliminativist and realist accounts of propositional attitudes. Building on Michel Seymour’s distinction between material and intentional components of meaning, I argue that the concepts of folk psychology postulated to explain others behaviour represent idealized types of mental states that correspond to neurophysiological types and serve to explain broad types of behaviour. The concepts of such folk psychology are material in Seymour’s sense; they belong to the public account of one’s behaviour and are perceived by external observers as causally efficacious. However, they do not square with the agent’s own account which is intentional: based in his own perceptions of his intents, the objects he posits into existence and idiosyncratic senses he attaches to the concepts of public language. Neither the material nor the intentional accounts are fully accurate. While the material folk psychology explanation represents an attempt to account for the agent’s behaviour in categories that are too general and broad, the agent’s own intentional account is but his own interpretation of the real neurophysiological causes of his behaviour. The true folk psychology theory of behaviour results from the process of social interaction between the agent and external observers that simultaneously brings about a modification of the idealized categories of material folk psychology, re-assessment of the agent’s own intentional account, and a corresponding adjustment in the underlying brain states.

Propositional attitude is a philosophical term for a psychological state, or a combination of a mental state with a specific propositional content.\(^1\) Propositional attitudes include the states of believing, desiring, hoping, fearing, despairing, enjoying, and many others. All of them characteristically express a particular attitude towards the propositions that constitute their meanings, or contents.\(^2\) Together the attitude component and the contents of propositional

---

\(^1\) An earlier version of this argument titled ‘Metaphysics and Semantics of Folk Psychology’ is soon to be published by Common Ground Publishing in *The International Journal of Humanities.*

\(^2\) These contents can be specified through the subordinate clauses following the conjunction *that:* I believe that the Earth is round; I desire that I be healthy; I fear that I may lose the bet, etc. Alternatively, these same contents may be rendered using the reified correlates of these states: I have the *belief* that the Earth is round; I have the *desire* that I be healthy; and I have the *fear* that I may lose the bet.
attitudes serve to individuate propositional attitudes themselves. Propositional attitudes constitute the linchpin of the everyday practices of description, understanding and prediction of behaviour. All of them are collectively known as commonsense psychology, folk psychology, belief-desire psychology or intentional psychology.

The precise meaning of "folk psychology" escapes attempts to define it. There are at least two reasons why sharp and uncontroversial definition of "folk psychology" is an impossible goal. In the first place, "folk psychology" stands for the practice of everyday explanation of behaviour, the practice which is largely uncodified and varies across people who use it. And secondly, folk psychology changes over time, both as a repertoire of practical skills and their theoretical underpinnings.

It seems beyond doubt that the developments within psychology itself, as well as philosophy of meaning, leave a mark on our concept and practice of folk psychology. In particular, the daily practices of attribution of psychological states became radically changed in the wake of psychoanalytic revelation of subconsciousness.¹ My own discussion of folk psychology in this paper will use the benchmark of the Cartesian model of the mind. In many respects, the Cartesian concept of the soul, radically distinct from and superior to the body, has laid the foundation for the self-understanding of the modern man.

To the pre-philosophical, naive mind, whose frame is set by the long-reigning Cartesian tradition, the existence of propositional attitudes and their contents does not seem in the least problematic. Nor does in fact the objectification of psychological states in the form of beliefs or desires, or the possessive relation I am alleged to stand in relation to them. Propositional attitudes constitute the multitude of everyday psychological life of each person. I confirm their existence anytime I reflect upon myself or my attitude towards the world. They are undeniably there as I try to verbalize the contents of my consciousness, indeed my beliefs, desires, fears and hopes. And, barring occasional inattention or sloppiness of my inner perception, they are by and large grasped accurately by the introspective "I".

These claims are far from obvious, to say the least, yet their questionable status is largely missed by the pre-critical thinking that still remains determined by the Cartesian worldview. In fact, prior to the systematic deconstruction of the mind initiated by the precursors of post-modernity, Nietzsche, Marks and Freud, the agent would have little reason to doubt that self-reflection can, in principle, provide a veridical account of the meanings of one's beliefs and desires. The traditional Cartesian model of the mind espouses jointly the theses of the First-Person View and the Transparency of the Mind.

¹For example, we are now more prone to look for the cause of one's behaviour to his suppressed beliefs or desires, the psychological states that the agent himself does not acknowledge and even denies having. It remains still an open question to what extent this and other changes improve the actual accuracy and efficacy of folk psychology as practice of explaining behaviour.
According to the former, "introspection provides the perspective from which to investigate the nature of the mental" yet it is due to the latter that introspection is veridical as "the meaning of a word or a sign is immediately present, unmediated, and transparent to the mind" (Cavell, 1993, p.17). The result is the claim of First-Person Authority which asserts the non-inferential and foundational knowledge of the contents of one's conscious acts.

While the post-modern hypotheses of subconsciousness have shaken self-knowledge of the determinants of our own behaviour, and consequently the Transparency of the Mind thesis, they did not affect directly the claim of First-Person Authority. Inasmuch as propositional attitudes partake in consciousness, their contents have continued, until recently, to be seen as fully known to the agent. Everybody was supposed to enjoy a privileged access to the contents of his or her mind, the latter believed to be filled with private mental objects mediating between the subject and the cognized object in the world. In Descartes' original version, this privileged access to the mind was additionally supported by a substantive distinction between the spiritual subject of consciousness and the material world.

The metaphysical and methodological questions are not entirely separate, and any answer to one of them will invariably weigh upon the other. In particular, if it can be proven that psychological states do not really exist, or are causally irrelevant to behavior, or if their contents are in fact quite different from what we believe they are, then folk psychology should be eliminated either as a theory or a practice or even both. Conversely, if folk psychology is indeed an implicit theory then it comes equipped with its own set of terms whose methodological status and metaphysical recognition will depend on the theory's performance. More specifically, folk psychological vocabulary could prove merely instrumental in case folk psychology does not describe the actual behavioral interactions but sets only the normative standards, and the terms of folk psychology would be outright denied any methodological use and theoretical existence if folk psychology failed altogether.¹

The jury on folk psychology is still out but the eventual verdict seems to depend on the three basic functions folk psychology is expected to perform. The first is public-referential and has to do with the external account of one’s behavior. By attributing propositional attitudes we are able to make sense of other people’s current behavior and predict their future actions. Or at least, this is what the realists and instrumentalists claim.

At the same time, belief-desire psychology is applied in a first-hand way, serving as the foundation of everyone’s self-account and rationalization of one’s behavior. This self-explanatory role need not individuate the same

¹In their distinct ways then, the controversy over propositional attitudes and the debate on folk psychology as such will converge on the same array of positions. They will extend over the whole spectrum of realism, instrumentalism and eliminativism; from the all-permissive realism according to which "propositional attitudes" are theoretical terms designating real psychological or neurophysiological objects, through instrumentalism which denies the references of folk psychology real existence but upholds their explanatory value, to finish at eliminativism which denies them theoretical worth and consequently existence as well.
propositional attitudes as public-referential since the account of another person’s behavior is based in propositional attitudes ascribed through a process of external rationalization while the self-ascription of beliefs or desires results from agents themselves making sense of their own behavior. The discord affects a wide swath of belief attributions, from mundane reports, like ‘Pierre believes that there is some beer in the fridge,’ to more theoretical rationalizations of the type ‘Pierre is obsessively checking the fridge for cleanliness.’ In either case, the external attribution of a propositional attitude (a belief in the first example and an obsessive-compulsive desire in the second) can be an attempt to rationalize Pierre’s action of opening the fridge. However, both attributions may be countered by Pierre’s own self-report which finds reasons for Pierre’s trip to the fridge in his empty stomach or a fastidious desire for new tastes. To further compound an account of Pierre’s behavior, it may turn out that Pierre maintains a peculiarly idiosyncratic concept of beer which everyone else identifies as a kind of bitter-tasting pop. If this is the case, then neither Pierre’s admission ‘I believe there is beer in the fridge’ nor do the observers’ attributions of beer-seeking intentions accurately explain his action. No matter what the true account of Pierre’s behavior is, such differences in attributions of propositional attitudes cannot be settled indisputably, and this seems to undermine any claim to real existence of propositional attitudes and even their instrumental usefulness.

The eliminativist conclusion is additionally fueled by the apparent failure of folk psychology’s integration within today’s scientific worldview. If commonsense psychology is to be seen as a science, its concepts should be accounted for in terms of neuroscience on which they supervene. (Kim, 1984) Sometimes this claim goes even further: it is also expected that any putative laws or regularities of folk psychology be reduced to the more basic laws of brain physics and biochemistry. However, the realization of such physicalist-reductionist role is unlikely. With the origin of materialist theories of the mind it has become exceedingly clear that posits of folk psychology resist attempts to reduce them to categories of neurophysiology, and the prospects of finding any bridge laws connecting folk psychological laws (suppose we could settle for any) with the laws of physics remain a just dream.

In this paper I shall sketch an account of propositional attitudes that attempts to take on the three roles of folk psychology. I should deem it a success if the concept of propositional attitude I put forth secures the satisfaction of the three roles or at least explains why some of them are conceptually impossible to satisfy and therefore inapplicable. I am going to base my project on Michel Seymour’s (1992) distinction between two interpretations of sentences that report on propositional attitudes of the kind

(1) JOHN BELIEVES THAT NEUTRINOS HAVE MASS.

On one of them, an atomic belief report, sentence (1) describes a particular atomic belief John tokens; on the other, sentence (1) is a general belief report which asserts the existence of a certain sentence type that is believed by John
and can be translated as ‘Neutrinos have mass’. The ambiguity results from
different interpretations of the ‘that’ in (1), and in particular its nominalizing
function. The first reading nominalizes the subordinate clause ‘Neutrinos have
mass’ as a quotation of a sentence type expressed through an atomic belief. It is
equivalent to

\[ (2) \quad \text{JOHN BELIEVES ‘NEUTRINOS HAVE MASS’}. \]

The second reading is modeled after Bertrand Russell’s theory of
descriptions and treats the expression ‘that neutrinos have mass’ as a pseudo
proper name. In result, sentence (1) becomes a general statement asserting that
John believes a sentence type that can be translated as “Neutrinos have mass.’
Its form is then:

\[ (3) \quad (\exists p) \left[ \text{believes (John, ‘p’)} \land \left( \text{‘p’ is translated as ‘Neutrinos have mass’} \right) \right] \]

To Seymour, ‘[t]he deep motivation for making such a distinction between
atomic and general belief reports is that we want to allow for the possibility of
reporting within the limits of our own language someone else’s belief with his
own understanding of linguistic meanings or his own principles of maximal
rationality.’ (p.183) This account forces a double way of individuating atomic
and general beliefs, with reference to language in which propositional attitudes
are expressed, and with reference to their actual contents. On the one hand,
Seymour seems to suggest that atomic belief attributions will be made using
the resources of the speaker’s own conceptual scheme-tending to his own
idiolect and merely quoting the uninterpreted contents of his mind in the form
of an encapsulated belief report. By contrast, general belief reports will simply
translate the speaker’s idiolectic expressions into the public language.
However, the reference to ‘maximal rationality’ suggests a different criterion:
belief reports may be individuated as atomic or general depending on how
closely the explanation of the speaker’s behavior matches the standards of
‘maximal rationality’ that is recognized by the external observer and which,
presumably, idealizes the rationality of belief attributions by society.

The distinction between atomic and general belief reports cuts across
another one. Within either category of belief reports Seymour identifies
intentional and material use of the attitudinal verb.\(^1\) To a large extent, the same
redundancy of criteria that blurs the atomic/general distinction afflicts
Seymour’s second classification of belief reports. And like with the
atomic/general classification of belief reports, equivocating on the criteria of
intentional and material attributions leads to the ambiguity in the taxonomy of
belief reports resulting from them.

On the one hand, the distinction rests with the purpose of the report: it is
meant to point to either the content of the reported belief or, rather, its

\(^1\)Like Seymour, I shall focus mostly on beliefs. A similar analysis extends however to all kinds
of propositional attitudes.
linguistic meaning. A report sentence that makes an intentional ascription of believing, desiring, hoping, etc. specifies the linguistic character of the sentence, which is the intended object of the propositional attitude. For example, intentional belief reports convey ‘a belief relation between the agent and a certain linguistic meaning under a mode of presentation that is a certain verbal form.’ (p.184) Intentional belief reports, so understood, are well suited to reflect the speaker’s own idiolect. By contrast, we ‘speak of a material use of the verb when the whole sentence describes a relation between an agent and a certain content and the sentence used no longer serves the purpose of specifying any mode of presentation at all for the agent.’ (ibid) When so presented, a belief report conveys the speaker’s contents expressed in the objective, referential terms of public language.

However, the other criterion of individuating intentional and material beliefs pulls in quite a different direction. ‘[W]e shall be tempted to make an intentional use of the verb “believes” in order to report the actual content of an intentional belief’ (p.185) with its ascription ‘founded either upon sincere assent or dispositions to assent sincerely to a sentence’ (ibid) while ‘the material use of the verb will ideally serve the purpose of reporting a functional teleological state of the agent.’ (ibid, my emphasis) On this criterion, the distinction seems to be founded on the cognitive and volitive states more than linguistic meaning. In the case of intentional belief reports propositional contents are attributed when the agent admits intending them. Material attributions, on the other hand, are made by the external observer on the grounds of some objectively recognized criteria of rationality.

To avoid ambiguity and keep the general/atomic and material/intentional distinctions separate I shall restrict the former to the sphere of meanings and make the latter centre around motivations. In other words, in the case of atomic belief reports (2) we merely describe the relation the agent holds with a particular sentence. John’s belief concerning the property of neutrinos is quoted literally, that is in terms of his own language and conceptual scheme, but regardless of his own motivational structure.

When a belief statement

(4) ‘Pierre believes that there is some beer in the fridge’

is represented as an atomic belief report

(5) ‘Pierre believes that “There is some beer in the fridge”,’

it states a fact of Pierre’s tokening of a certain mental belief-statement that, to Pierre, means ‘There is some beer in the fridge.’ Such a belief attribution does not account for the fact that Pierre’s concept of ‘beer’ is confused and what he takes for beer is in fact a bitter-tasting pop. Neither does it allow for a possibility that Pierre’s belief, or his fridge-searching behavior which seems consistent with such a belief, may actually point to a different volitive and
motivational structure on his part, for example a desire to deceive the observers into thinking he is looking for beer.

By contrast, through a general belief report (3) we assert the existence of the agent’s belief and provide a translation of its content in terms of our own conceptual scheme. A belief statement (4) interpreted as a general belief report

\[(\exists p) \] [believes (Pierre, ‘p’)) & (‘p’ is translated as ‘There is some bitter-tasting pop in the fridge’)]

allows us to understand the contents of Pierre’s belief using the terms of our own language.

The intentional/material distinction is, on my proposal, tuned in to issue of the correct determination of the agent’s intent, in the psychological sense of the will and motivation. Being focused on the agent’s own ‘take’ on his beliefs, intentional ascriptions attend then to the intentional states of the agent, the way he recognizes and admits having them The ascription of belief is based on the agent’s sincere assent to a sentence, or a sincere disposition to assent, and is characterized by a first-person authority. External assessments of the agent’s apparent motivations play no role in the determination of content of his beliefs, and the standards of objective rationality of behavior do not apply either.

The material way of ascription, on the other hand, attends to the content of the attributed belief-sentence, as determined by the external observer and the intersubjectively recognized rules rational reasoning. The report sentence assigns propositional contents and attitudes which need not be confirmed by the agent himself. Rather than report on the agent’s avowed intentions, it prescribes content as a functional teleological state of the agent. Material ascription of beliefs stipulates therefore contents that typically are not intended or apprehended by the agent and may be even unconscious. Their attribution results from a purely external description which attempts to make the best sense of his behavior.

The two classifications of belief reports result in four different types of belief ascriptions. Atomic intentional belief sentences, like

\[(7) \] JOHN BELIEVES, ‘NEUTRINOS HAVE MASS’,

report directly John’s intentional beliefs, the way he is disposed to assert them and using his own concepts of neutrinos and mass.

On the other hand, atomic material belief reports, like

\[(8) \] JOHN BELIEVES,M ‘NEUTRINOS HAVE MASS’,

are used to relay contents which are directly inferred from John’s behavior, rather than confirmed by him, but which are still presented in terms of his own idiolect.

Now a general intentional belief sentence

\[(9) \] (\exists p) [BELIEVES, (JOHN, ‘p’)) & (‘p’ IS TRANSLATED AS ‘NEUTRINOS HAVE MASS’)]
translates the content of John’s belief, the way he recognizes it, into the language of the reporting audience, while a general material belief report

\[(\exists p) \text{[believes}_m \text{(John, ‘p’)} \& \text{‘p’ is translated as ‘Neutrinos have mass’}]\]

offers a public language description of the content of the belief that is externally attributed to the agent.

This much shall do for an account of Seymour’s article. His own theory of propositional attitudes is meant to deal with some of the traditional problems associated with a quotational approach, like ascription of beliefs to non-linguistic creatures or failure of interchangeability between synonyms. For me, however, the typology of propositional attitudes ascription I derive from Seymour will serve a different purpose; in fact, two distinct but related goals. First, it will help account for the diverse and seemingly discordant ways of attributing beliefs and desires in folk psychology. Second, it will allow me to show how we can make folk psychology satisfy two of its roles and how we can explain away the third one.

Let us start with a rather trivial observation. Particular types of belief reports differ not just in their grammatical form or relation to the agent's idiolect and rules of rationality. In fact, they point to different propositions in the sense of logical meaning of a belief-sentence attributed to the agent. A proposition in this sense can be interpreted as a set of possible worlds in which it is true.\(^1\) With every belief ascription we end up constituting a distinct notional world that conforms either to the agent's conceptual scheme--with or without his own rationalizations--or to the linguistic standards of the broader community, again with the agent's own motivational account accepted or overruled. Seymour's intentional atomic belief reports convey the propositional attitudes of the agent the way he discovers them introspectively and in his own idiolect. They reflect then the agent's own rationalization of his behavior and convey his story in terms of his own, possibly unique or idiosyncratic, linguistic idioms. His individual senses point to their own referential counterparts, the objects which they denote and which, however fictitious, populate his notional world.

Material atomic belief reports are still verbalized using the agent's own conceptual resources; however they no longer attribute propositional attitudes asserted by the agent himself. Rather, their ascription results from an external assessment which sticks him with those mental states that make the best teleological sense of his actions. The agent's own canons of rationality, including his own individual value choices, tastes, and even propensity to errors in reasoning, give way to the standard, socially determined rules of decision making. The objects of his notional world remain the same as in the case of atomic intentional belief ascription. Material assessment of his action

\(^1\)Obviously, such possible world semantics is just one a few different conceptualizations of proposition; however, it is most natural in this context—after all, Seymour's typology of beliefs is based in Kaplan's possible worlds semantics analysis of indirect speech. (Kaplan, 1989)
A Physicalist and Explanatory Concept of Propositional Attitudes

does not affect the references of his idiolectic senses; they are just positioned in a different motivational and volitive structure.\(^1\)

With general belief attributions one no longer quotes the agent's propositional contents; the prism of the notional world populated by references of his own characters is replaced with a more objective reality determined by the public language. The translation merely explains what the agent understands by his propositions; it shows how the notional world of his linguistic meanings and their referential counterparts can be related to the world of the public language. Intentional general belief reports still preserve the agent's own rationalizations—it just translates his story into the language of the broader linguistic community. Yet with material general belief sentences the agent's account is completely lost to the external point of view, with respect to both its linguistic form (as a translation) and the attributed contents of his beliefs or desires.

With all the types of belief reports interpreted in terms of possible worlds semantics, their relevance to folk psychology becomes clear. Notional worlds determined by belief reports point to distinct spaces of linguo-psychological accounts of behavior, drawn from the perspective of the agent himself or that of the external spectator. The most subjective attribution in the form of an intentional atomic belief represents the agent own rationalization. As it explains his behavior through his own principles of rationality and in reference to the objects of his own notional word, the intentional atomic belief account satisfies the self-explanatory function of folk psychology.

On the other end of the spectrum of belief attribution, the material general belief reports present an explanation of the agent's actions in terms of the public language and external teleological rationality. The agent's own account tends to be ignored; though the attribution translates the idioms of the agent's idiolect into the public language, and so accounts for the objective references of his notional world concepts, his own psychological explanation of behavior is irrelevant if it conflicts with objective rules of teleological rationality. This focus on referential accuracy and prediction of behavior makes general material belief reports suitable for the public-referential function of folk psychology.

As I noted at the beginning, one of the biggest issues in folk psychology is the discrepancy between one's own account of his behavior and the public account of it. The unavoidable difference in attributing propositional attitudes threatens the status of folk psychology as a reliable tool of understanding oneself and others. My analysis so far traces this difference to the different vocabularies and methods of ascribing mental states. While the general belief reports are based in the public language and the world of commonsense experience as its denotation, the atomic beliefs are founded in the agent's idiolect and the correlate notional world it intends. Differences between the languages translate into different world of intended references and the accounts

\(^1\)If I neglect to help a 'geek' out of spite or just an oversight does not affect my peculiar understanding of the word 'geek' as a person who wears Don Cherry style tuxedos. One or the other interpretation means only a different attitude I have toward geeks.
of behavior that talk past each other. Going back to an example offered before, Pierre’s explanation of his beer-seeking behavior remains incomprehensible to the external audience until his idiosyncratic understanding of ‘beer’ is uncovered.

The intentional/material distinction complements this picture with an examination of the psychological account of behavior. The material belief attributions result from an external assessment and stick to the rules of teleological rationality. The intentional belief reports, on the other hand, stem from introspective self-analysis. The propositional attitudes they deliver are asserted by the agent himself; they represent the agent's own modes of psychological presentation for the objects of his notional world. The attitudes I find in self-reflection may not be the same as those attributed to me by external observers. Even when the linguistic differences between general and atomic attributions are fixed, the psychological accounts of behavior are dramatically divergent.

Now the self-explanatory and public-referential functions of folk psychology are satisfied by two radically different methods of attributing propositional attitudes. Beliefs or desires individuated through atomic intentional reports fulfill the self-explanatory role; however, the public-referential function is best performed by propositional attitudes determined through the general material report sentences. To save folk psychology as a method of explanation and prediction of behavior we need to reconcile the differences in both linguistic references and psychological senses. The sought-after account of folk psychology will emerge then as a final product of a two-stage process. In the first place, we need to reconcile the differences in the language of psychological evaluation. This means a translation from atomic intentional beliefs to these same propositional attitudes expressed in the general intentional form. The resulting accounts of behavior are spelled out in terms of the same language; consequently they point out to the same objective references. However, the accounts are still not located within the same notional world. The actual psychological accounts differ; the agent tends to fill his account with different motivations and attitudes than the external observer. To risk a pun, the psychological senses of their assessments are worlds (notional) apart.

Here is where the final reconciliation will take place. The agent's own account of behavior will clash directly with the account determined through an external observation. Agent-avowed motivations will have to withstand the evidence available to the third party observers. If the agent's own attributions cannot be squared with propositional attitudes inferred from his behavior, his intentional belief reports will have to be modified. Conversely, if the assumptions of teleological rationality are denied by the agent and his own principles of behavior without any obvious inconsistency, then agent-determined senses will be adopted as the basis for the folk psychological account of his action.

In most cases modifications will likely come from both sides. The process of establishing the concept of content for folk psychology is in fact a
hermeneutic dialogue where both positions grow and change as they trade reasons and evidence. The eventual contents are sealed by the requirement of maximum consistency. In the end, when the agent and the external observers agree on his propositional attitudes, the self-explanatory and public-referential functions of folk-psychology are concurrently satisfied.

So far my discussion has focused on the first two functions of psychological content, self-explanatory and public-referential, and left out the third, physicalist-reductionist. For its satisfaction critics of folk psychology have traditionally required two tasks: that we present an account of how the concept of content in folk psychology is reducible to brain states and how the laws of folk psychology can be derived from neurophysiology. Now the latter task is today widely discredited. It is unfair to expect of folk psychology meeting methodological standards that no macroscopic science can fulfill. No one blames geology or economy for not submitting their laws or regularities to microphysics. The wider the ontological gap between the sciences' levels of description, the more implausible such a demand becomes. The fact that the bridge laws connecting the regularities of a higher level, macroscopic science with the laws of physics governing brain events just don't exist in no way detracts from the validity of the macroscopic science. If geology and economics subsist as legitimate sciences, so should folk psychology.

The first task is more difficult, and my account of folk psychological content will turn out to force a seemingly paradoxical solution. Even if the regularities of folk psychology cannot be reduced to the laws of physics, grounding folk psychology in physics requires that its natural kinds, that is posits invoked in a folk psychological account, could be shown to supervene on brain events. Again, it has become a commonplace, at least since the nineties, that the individuation of natural kinds does not advert to the criteria of completeness of microcausal account or supervenience of natural kinds upon physics. In any science, intentional or non-intentional, natural kinds are always individuated in the way that maximizes their explanatory and descriptive powers.

However, with folk psychology fending off the challenges of eliminativism and instrumentalism, this becomes a moot point. To convince the critics that the propositional attitudes of folk psychology are real one needs to show that they satisfy the principle of psychological autonomy

that the properties and relations invoked in an explanatory psychological theory must be supervenient upon the current, internal physical properties and relations of organisms (i.e., just those properties that an organism shares with all of its replicas).

(Stich 1978, p.575)

So far, the argument of this paper has proved that neither intentional nor material method of attribution of propositional attitudes is likely to deliver a real belief of the agent. Consequently, there is nothing in the brain the intentional or material ascriptions could supervene on. The failure of
neurophysiological individuation of folk psychology's natural kinds is then far from surprising. It is also quite understandable why folk psychology has a mixed record explaining and predicting behavior.

However, the hermeneutic content resulting from the dialogue between the agent and the external observers is one that the agent does acquire, self-reflectively owns up to, and which proves to account for the external evidence from the third-party observation. It is a truly occurrent belief, and as a real belief of the agent, it must point back to some neurophysiological realization—to deny this would push one toward a metaphysical dualist theory of the mind. As a belief which both the agent and external observer alike accept as the true account of behavior, it must stand in a verified causal relation to other beliefs and the behavior of the agent. As causally dependent on behavior, it supervenes on it. Now that agent behavior itself supervenes on his brain events is an assumption that the materialist critic of folk psychology dares not deny. It is also clear that the relation of supervenience is transitive. If so, then the hermeneutic content, too, must supervene on neurophysiology.

The hermeneutic account of folk psychology leads eventually to the concept of content that satisfies all three functions expected of it. Beliefs and desires resulting from a reconciliation of intentional and material atomic reports are acceptable to the agent and the external observer, and supervene on neurophysiology. Since the concept of supervenience assumes the existence of nomological connections between the natural kinds of both theories, my theory asserts that folk psychology regularities must, after all, be reducible to laws of physics. I must admit, however, this is a hollow victory. The contents attributed to the agent may be just too fine-grained for us to formulate any law or generalization over them. The hermeneutic process of their individuation required both translation from the agent's idiolect and accommodating initially conflicting evidence from introspection and external observation. If not holistic altogether, then the hermeneutic content gives rise to propositional attitudes that are too unique to enter any regularities we could identify. It is unlikely we would be able even to typify them. The same can be expected of the brain events on which they supervene. As individual as the beliefs or desires they subvene, they won't allow us to discover any types or laws over them. We should be satisfied just knowing that folk psychology is in principle reducible to neuroscience; in practice however we will never know about it.

References

Many philosophers now writing on the nature of human morality follow a trend of scrutinizing the latter by a naturalistic or scientific approach. These philosophers refer to a novel picture of morality that advocates ‘the new synthesis’ (Haidt, 2007) based on biologically evolved and culturally sensitive brain systems.

This new conception of a relatively old field of study which is morality becomes now known as moral psychology. As claimed by Haidt, its explosion was much predicted by E.O Wilson's account of the origins of morality understood as a cross-disciplinary merge of ethical behaviorism and the cognitive-developmental theory on the one hand, and a research on the hypothalamic-limbic system, on the other. Yet, as ethical behaviorism faded with philosophical behaviorism, so did Kolberger's trust in hypothalamus, the last one found to play little role in moral judgment. Being focused on conscious moral reasoning, ‘Kohlbergian psychology forged its interdisciplinary links with philosophy and education, rather than biology as Wilson had hoped.’(Haidt, 2007)

It is unquestionable that Wilson has misplaced his hope for the merge of the three theories but, as claimed by Haidt, his mistake was only because he ‘got the ingredients wrong.’ His overall picture was right though. Ethics has not and will not merge the two psychological theories of Kohlberg's days with research on the hypothalamic-limbic system but as a cross-disciplinary study it is still to be based on ‘a set of ingredients.’ In fact, the process began somewhere in 1990s and although the list of ingredients vary from social psychology to neuroscience and evolutionary theory, the new synthesis grows rapidly in the last few years. In the meantime, E. O. Wilson got his way to moral psychology pantheon whose visionaries recognize that the hope for our understanding of morality does not lie in traditionally approached academic fields of descriptive versus normative ethics or meta-ethics. It is rather the development of our knowledge of morality spelled out in terms of the proper functioning of circuits in the prefrontal cortex or in terms of products of natural selection that is to revolutionize our moral knowledge.

So far, the prospects of achieving good results from ‘the new synthesis's’ implementation are sufficiently robust to warrant this philosophically new trend. As moral psychology is progressing in several directions, we are finally in a position to find at least some neural underpinnings or correlates of moral sensitivity or we are able to explain brain mechanisms that underlie what philosophers have called moral judgments. We may reasonably expect that
someday so called neuroethics will provide us with explanation for the rest of traditional ethical concepts like moral value, moral character, moral development or moral obligation.

However, there is also a growing consensus that moral psychology, in as much as it remains committed to the empirical method, questions human powers of reasoning and reinforces the intuitionist perspective on morals. Or, at best, being focused on the facts about the human brain, moral psychology is essentially concerned with actual cognition, reasoning or behavior. As such, it ignores the conceptual work of normatively oriented philosophers concerned with ideal standards and the agency of human beings. What follows is that one of the most interesting yet troubling features of moral psychology and in particular neuroethics is that it blurs a sharp distinction between facts and values or the ‘is’ and ‘ought to’ of traditional ethics. As a rich and exciting field of research, moral psychology provides us with what seems to be a very efficient method of understanding and explaining human behavior; it allows us to explain ourselves as well as others and it may allow us to make fairly reliable predictions about the morally significant behavior of ourselves and others. Yet, at the same time taking a form of a causal theory, moral psychology has no bearing on the status of non-causal explanations by reason, otherwise known as moral justifications.

Where do we go from here then? We could argue that the history of normative ethics provides a prima facie reason for the acceptance of moral psychology. In this view, our powers of reason are only improved abilities to articulate what is intuitively available. They are causally generated by moral intuitions and as such can be equated with them. In G. A. Cohen's vocabulary, they are as 'fact-sensitive' as intuitions themselves and any attempt at generating 'fact-insensitive’, normative principles is not valid. In this sense, our moral intuitions have both the privileged epistemic and ontological status, and any moral theory being descriptive or normative serves one purpose of discovering more or less hard-wired principles. However, the nature of these principles considered from the perspective of moral psychology remains the same. This reductionist account, although not necessarily fueled by neuroscientists and philosophers alike, provides a single unified account of moral world nonetheless and one could argue that this is perhaps the biggest-yet unintended- importance of the questions investigated by moral psychology. In its outcome, artificial boundaries between descriptive and normative subdivisions of ethics are crossed and we are getting a simple typology as a result. However if identification of the object of study is the first step in any scientific or philosophical endeavor, the force behind moral psychology becomes obvious.

But does moral psychology really imply that philosophers are not obliged any longer to do conceptual work defining morally good actions, or differentiating between what is rational or irrational, intentional or unintentional? No. In the first place, articulating the relations between these concepts and giving accounts of the difference between them seems necessary in order to describe and interpret the facts of empirical moral psychology.
Secondly, even if in terms of empirical research we recognize the tendency to base moral judgments on moral intuitions, we still would like to think ‘that we have some capacity to think through moral issues, to weigh up alternative outcomes and make a call on what is right and wrong.’ (Jones, 2008) Either way, the argument sounds circular or counter-intuitive.

We may find it important and interesting then that some recent experimental studies point that our moral intuitions do not have the privileged status they were previously thought to have and ‘neuroscience gives some cause for optimism’(Jones, 2008) that our judgments are not simple intuitive perceptions of what is morally permissible or not. In other words, it is the very empirical research that suggests that there is a difference between intuitive—hence descriptive in nature claims, and our moral claims accounted for in terms of rationality and normativity alike. There is no primacy of intuitive morals and, philosophically speaking we should not prioritize intuitions because empirical study poses a serious challenge here. If so, and even if the relation between intuiting morals and reasoning about morals might occur fairly complex and problematic, Mackie's argument that moral norms and values are too 'queer' to be part of the furniture of the world does not stand the test of empirical findings of cognitive approach. (Mackie, 1977)

To summarize briefly, cognitive scientists, mostly Joshua Greene but also Antonio Damasio, used functional magnetic resonance imagining to map the brain when its subjects were trying to solve traditional moral dilemma. In particular, Greene focused on what he called impersonal moral dilemmas- like the well known Trolley Problem or a case of receiving a letter from a reputable international aid organization asking to make a donation of two hundred dollars, and personal moral dilemma-like the Footbridge Problem or, adapted from Unger the case of 'driving along a country road when you hear a plea for help coming from some roadside bushes.' (Greene, 2003) The first two cases ('trolley' and 'donations') as explained by Greene are impersonal because they involve the moral violation which does not occur in an 'up-close-and-personal' manner. In the trolley case, you simply need to switch the trolley sending it onto another track on which just one person is walking, while the trolley originally is heading down a rail track with five hikers on it. In the case of donation you are asked to assist people in far-off countries and even if their needs are severe, the issue does not occur to you in an 'up-close-and-personal' way.

Yet, the dilemma with driving along a country road and the Footbridge dilemma are personal. While hearing the plea ‘you pull over and encounter a man whose legs are covered with blood. The man explains that he had an accident while hiking and asks you to take him to a nearby hospital. Your initial inclination is to help this man, who will probably lose his leg if he does not get to the hospital soon. However, if you give this man a lift, his blood will ruin the leather upholstery of your car.’ (Greene, 2003) As such, it becomes your personal dilemma whether you should leave this man by the side of the road in order to preserve you car seats clean. Similarly, in the Footbridge Problem you are engaged in a personal dilemma that invokes up-close and personal violence
of pushing a man in front of the trolley. The Footbridge Problem involves the same five hikers ahead of the trolley, but now there is no switch available and the only way to save the hikers is to put a heavy weight in front of the trolley. Since the only weighty object nearby is a large man standing on the footbridge with you, you need to make a personal decision whether to push the man or not.

Conducting the study and mapping up what's going on in the brain when people reflect on these sorts of personal and impersonal dilemmas, Greene finds out that 'judgments in response to personal moral dilemmas, compared with impersonal ones, involved greater activity in brain areas that are associated with emotion and social cognition'. He argues that the personal dilemmas indicate that 'moral judgment is often an intuitive, emotional matter' or otherwise an effortless, cognitive process accompanied by perceptual phenomenology. Like we can effortlessly determine whether a given face is male or female, we effortlessly 'see' rightness and wrongness and form moral judgments. This is because we have evolved mechanisms for making quick, emotion-based social judgments since our social lives favor such capacities. Natural selection has equipped us with mechanisms for making emotion-based moral judgments quite useful while searching for the most suitable mates, for example and now assists us in finding answers to personal dilemmas.

So much about Greene's evolutionary account of the latter. On the other hand, functional magnetic resonance imaging of the brain, as performed by him on people thinking through impersonal moral dilemma, showed increased activity in the brain regions in the prefrontal cortex that are associated with deliberative reasoning and cognitive control. Similar results were confirmed by Antonio Damasio's study of emotional circuits in the prefrontal cortex. Damasio found that patients who sustained damage to certain regions of the prefrontal cortex did not lose their cognitive capacities by most measures, including IQ and explicit knowledge of right and wrong. However, because of the damage these patients exhibited a massive deficit in their emotional skills which crippled their judgments and decision-making processes. In other words, the patients lost their intuitive abilities but could still deliberate about what's good and right. Their moral intuitions-fast, automatic feelings of what is good and what is bad- were gone due to the injury, but they could still reason in a controlled way. This ability to transform the information to reach a moral judgment devoid of moral intuition was equally supported by Greene's further research. When his subjects were asked difficult questions like: Would you smother your baby if it was the only way to keep her from crying and giving away your hiding place to the enemy soldiers looking for you, who would then kill the whole group of you hiding in the basement?, they were slow to respond. Yet, their brains overcame the lack of initial intuitive response. Their brains exhibited increased activity in the anterior cingulate cortex, the areas found by Green as responding to internal conflicts. Some subjects' exhibited activity in the prefrontal cortex suggested additional processing and overriding the initial flash of horror.
To conclude, either way, the impaired brain, being damaged or lost due to the lack of intuitions, is still in a position to counterbalance the emotional void. ‘Reason is clearly not impotent in the moral domain. We can reflect on our moral positions and with a bit effort, potentially revise them.’ (Jones, 2008) This brings hope to rationality’s precepts spelled out in terms of moral principles or normative judgments discovered a priori. In a sense, empirical research accounts for non-empirical character of some moral claims.

Contemporary moral cognitivism’s conclusion that we don’t have to use intuitions as a basis for our moral claims provides a reason to adopt constructivism as a meta-ethical view. To that end, Jakob Elsner argues, for example, that meta-ethics leads to the understanding that empirical moral psychology is not a constraint on moral theory building but rather the material out of which moral theories are built. With the structural description in place, Elsner claims that ‘we cannot and should not do moral theory in order to revise moral practices from the outside; any revision of our moral practices must come from the inside, starting with these moral practices themselves.’ In like manner, Marc Hauser and John Mikhail offer a structural description of morality viewed as a system which uses ‘conversion rules’ to break down observed or imagined behavior into a meaningful set of actions which is then used to create a 'structural description' of the events. This ‘structural description’ captures not only the causal and temporal sequence of what happened and when it happened, but also intentional aspects of events like for example was the outcome of an action intended as a means or a side effect?

Based on his structural model, Hauser speculates further that the causal and intentional aspects of events can be compared with a database of unconscious rules, such as ‘harm intended as a means to an end is morally worse than equivalent harm foreseen as the side-effect of a goal.’(Jones, 2008) Accordingly, if the events involve harm caused as a means to the greater good (like suggested in the Trolley or Footbridge dilemmas) then it is more probable that our moral faculty will generate the judgment of moral impermissibility. If the events involve harm as a side effect then a judgment of permissibility is more likely to be generated. Quite interestingly, in this view the relationship between moral judgments as to what is morally right and wrong and emotions becomes reversed; it is emotions which arise as a by-product of unconsciously reached judgments and not vice versa as claimed by Green with regard to impersonal dilemmas.

Hauser argues further that the nature of the principles may be empirically tested; if universal, so present in all cultures, they should be understood as innate parts of the moral faculty. Yet, since cultures vary in their understanding of how much harm is tolerable, even as a means to the greater good, Hauser’s expectations indeed await a thorough empirical test, as mildly criticized by Jones. Jones seems to be right emphasizing that ‘a full account of our moral psychology will also have to explain the variations in human intuitions’ and ‘the problem of how and why societal norms of moral conduct change over time’ (Jones, 2008), to this extent Hauser’s account is not complete yet. What
interest me the most though is that Elsner and Hauser are both willing to grant moral judgments the meta-ethical status of a structure.

In my understanding the concept of a structure is the most suitable to account for a normative theory of traditional ethics without leaving aside recent attempts to naturalize ethics or think of it as ‘the new synthesis.’ If I understand Elsner's and Hauser's argument accurately, I also believe that traditional ethical theory constitutes the structure and its objects (in particular moral judgments) have no properties independently of those relating them to other objects of the same structure (in particular, emotions or intuitions alike). I propose however to view structural nature of ethical norms as similar to the objects of modern mathematics. In modern understanding, the objects of mathematics like natural numbers, groups, sets, spaces, and so on are thought to be determined by the relations which constitute the structure to which they belong. It has been argued that what is essential to them is that they have no properties independently of those relating them to other objects of the same structure. As there is indeterminacy with respect to just what object is picked out by any given number (is it a table, a chair or a mug as quoted by Hilbert in the context of geometrical objects), there is no determinacy with regard to other mathematical objects either. Structurally and relationally understood, mathematical objects have no physical referents and as such they cannot be individually identifiable in the way perceptual objects are. Perceptual objects have some single referent by which they can be picked out but quite contrary mathematical objects are ‘incomplete’ or undetermined in the way perceptual objects are not.

To expand a bit, for perceptual objects it seems quite reasonable to suppose that there is some sort of causal relation between the objects and the sensory apparatus of the person who sees them. It is presumably at least a necessary condition that there be such a relation in order to have a perception of the object. In the case of mathematical objects, however, it is not reasonable to suppose that there could be anything like causal relations. Abstract by their nature, they are not to be represented by any concrete spatial instances like, to repeat, tables, chairs or mugs. They do not involve any directedness to perceptual objects and they have no representing content at all. As such they do not have any causal attributes and amount to abstractions which consist of purely relational constructs. This distinctive feature of modern mathematics in virtue of which it stands as a paradigmatically structural and relational discipline can be associated with Hilbert's program for foundations of mathematics, but also rightly assigned to Bourbaki's or Godel's views.

Now, since my basic interest goes with the status of moral normative judgments, I maintain that like the objects of modern mathematics, they find their ontological and epistemological credentials in the concept of a structure. Norms, like the mathematical objects do not codify the particular rules of procedure and don't require to think of anything else in order to individuate them. Understood as relations, they have no 'hard-wired' existence on their own and are only possible to be seen from a frame of the whole structure of morality they are part of. Yet, themselves they are radically free of any physical
Structural Norms in Light of Neuroscience

interpretation. Like mathematical objects are not restricted to spatiality, moral norms are not restricted to intuitions or neurons, and represent some generalization and abstraction from them.

Contrary to norms, the status of moral intuitions is non-relational. Hypothetically, Greene admitted it himself while acknowledging that ‘an interesting feature of many intuitive, effortless cognitive processes are accompanied by a perceptual phenomenology.’ Although now, I believe, his analogy between seeing a person's maleness or femaleness and 'seeing' rightness and wrongness is more meaningful. Since the concept of intuition refers on his account to something that 'takes room in the spatio-temporal world, it provides a handle for all kinds of empirical discoveries'. Yet, as once noted by Nagel, another concept might clearly exclude the possibility that what it designates has certain features. The concept of structural norms seems to be such a case. To clarify, according to Nagel, ‘we do not need a scientific investigation to be certain that the number 379 does not have parents. There are various other things that we can come to know about the number 379 only by mathematical or empirical investigation, such as what its factors are, or whether it is greater than the population of Chugwater, Wyoming, but we know that it does not have parents just by knowing that number.’ (Nagel, 1998) I claim that by analogy, knowing that moral norms are relational we also know that we cannot expect them to be identical with any forms of 'perceptual phenomenology'. Yet, there might be many things we can come to know about judgments only by empirical investigation. To proceed with the analogy further, Nagel says that ‘if someone rebuked us for being closed-minded, because we can't predict in advance what future scientific research might turn up about the biological origins of numbers, he would not be offering a serious ground for doubt.’ By the same token, if someone rebuked us for being in the grip of meta-ethical or, in general, any non-cognitive approach, instead of adopting its competitor like moral psychology or neuroethics, she would not be offering any reasonable doubt either.

To conclude, I think that in concentrating on norms moral theorists have chosen to look at objects from a certain frame of reference in which they are individually identifiable yet not finite. Unbeknownst to them, these philosophers have chosen a mode of thinking that enabled them to reason and calculate with confidence beyond the scope of 'perceptual phenomenology'. Tacitly assuming the concept of the moral structure, they accounted for morality model that eliminates questions about the extent to which normative ethics satisfies the ontological commitment imposed by evidence from neuroscience. I find this model quite satisfying and claim that structurally understood norms are directly squared with the concept of moral intuition endorsed by a perceptual phenomenology of neuroscientific experiments. Morality does not have to be a product of cultural evolution after all and our moral truths do not have to be viewed as the contingencies of our evolved minds. Building on the material of which our moral theories are built, we may discover structural and relational moral norms.
An Anthology of Philosophical Studies

References

PART 3
Ethics and Value Theory
The discussion in contemporary normative ethics – at least in the tradition of analytic philosophy – is characterized by different and often rival approaches. Some of the most relevant normative positions in the debate are preference utilitarianism, contractualism, and contractarianism. After the contributions of John C. Harsanyi in the 1950s and 1960s, and later on of Richard Hare and Peter Singer, utilitarianism experienced a true renaissance. In 1971, John Rawls’s *A Theory of Justice* fully reintroduced into the discussion contractualism, thus renewing the whole field of ethical and political theory. Finally, Robert Nozick and John L. Mackie, also in the 1970s, laid the basis for contractarianism – an alternative version of the approach centered on a hypothetical contract that combines aspects from Hobbes’ *Leviathan* and, to a certain extent, Hume’s *Treatise*.

When moral philosophers construct their own theories, they characteristically pay attention mainly to meta-ethical and normative aspects. Consequently, criticisms to rival normative approaches focus on the meta-ethical starting point they have established or the normative dimensions they include or exclude in their systems. For example, the different types of contractarianism have been criticized because they generally cannot include animals – or even children – as part of the ‘moral community’; or act utilitarianism has been condemned because it may allow – at least in its less refined versions – the violation of basic individual rights in order to achieve an increase of the total sum of happiness. (Sen & Williams 1982)

---

1 For an overview, see Singer (1991, Part IV). For the difference between ‘contractualism’ and ‘contractarianism,’ see Darwall (2003).

2 Following the more or less well-established convention, I distinguish among meta-ethics, normative ethics, and applied ethics. Meta-ethics deals especially with the question concerning the justification of our moral judgments. The discussions between realism and antirealism, and between foundationalism and coherentism are among its central issues. Normative ethics purports to formulate a comprehensive and rational system of fundamental normative principles and norms intended to regulate individual action and social institutions. In general, the philosopher comes to the formulation of a normative system once she had previously defined her meta-ethical position. Finally, applied ethics can be seen as a particular domain of normative ethics. Its point is to establish concrete norms and criteria for specific situations, like in the case of medical practice.

In this paper, I will maintain that ethical theories base on sociological assumptions, i.e., on a particular conceptualization of human action. Moreover, these assumptions play a key role in the ‘economy’ of the normative theory. Nevertheless, that ‘sociological dimension’ remains mainly implicit or does not receive an adequate treatment on the part of the moral philosopher. If I am right, the consequence is apparent: Ethicists should pay more attention to the sociological aspects of their theories. The focus on – and the analysis of – those sociological assumptions will give them key insights for the construction and evaluation of their normative positions. In other words, I am addressing the need to systematically concentrate on the action-theoretical dimension which underlies every normative theory.¹

I will start by distinguishing the approach I am advocating from other possible types of relationship between sociology and ethics. Then, I will present a framework with the most relevant kinds of action-theoretical assumptions to be found at the core of ethical thinking. After that, I will briefly concentrate on a case study: after presenting Mackie’s contractarianism, I will underline its sociological assumptions and show which role they play in the structure of his normative system. Finally, I will conclude discussing how a systematical analysis on action-theoretical aspects can help to evaluate any given ethical theory.

Sociology and Ethics

To restate my point: I am arguing for the recognition and intensification of a particular kind of relationship between sociology and ethics. When constructing a normative theory, philosophers make action-theoretical assumptions. Clearly, there are other types of relationships that sociology and ethics may have. I would like to mention three other ways in which ethics can benefit from sociology.

Firstly, empirical sociological studies are a primary source of information about social issues that may require an ethical reflection. For example, the picture that sociology can provide about poverty or social stratification in a given society or worldwide is a decisive starting point for a reflection conducting to a theory of justice.² Secondly, the sociologist can indicate the ways through which a certain ethical goal can be effectively achieved. In this sense, the sociologist as a ‘social engineer’ shows the means to be employed to

¹I am not disregarding the role that the analysis of normative and meta-ethical aspects plays in the construction of moral theories – the point is, nevertheless, to underline the importance of that additional but so far mainly overlooked approach to moral theorizing. (Breuer 2007: 1-10; 161-168)
²Volker H. Schmidt (2000) argues that sociology plays an essential role in applied ethics, for instance in questions concerning the just allocation of scarce resources. The social scientist contributes decisively to the work of the practical philosopher showing the institutional and technical complexity which is constitutive of specific situations like in the case of organ donation and transplantation. Otherwise ethical considerations become ‘vague’ and ‘unsubstantial.’
reach the ends specified by a normative theory. (This is especially relevant for
the so-called consequentialist approaches in ethics, even though most ethicists
nowadays – even the ‘principle-centered’ ones – take into consideration the
practical consequences of their moral theories.) Finally, the social scientist can,
resorting to her expertise, properly assess the point up to which an ethical
theory can be realized or not. A moral theory can be an exquisite product of our
imagination, but it will be just a mental exercise if it is not realizable.¹

These three ways in which ethical reflection can benefit from sociology are
more or less well known and their relevance is generally recognized. Nevertheless, they are relationships of an ‘extrinsic’ kind. The particular way
of cooperation that I am pointing out is of a different and more radical sort. I
maintain that philosophers, at the very moment of constructing a normative
system, make use of a particular conception of the individual as a social actor.
That conception becomes the ‘foundation’ upon which ethicists’ normative
‘intuitions’ take shape. Sociology enters thus into the very core of ethical
theorizing.

It is convenient to make it clear that the position I am advancing in no way
seeks inappropriately to draw normative conclusions from social facts or
anthropological traits – thus overlooking the impossibility of deriving ought-
sentences from is-sentences.² What I am saying is this: in constructing an
ethical theory, philosophers forge their moral ‘intuitions’ with the help of a
certain conception of human action which necessarily entails sociological
assumptions. It is not about denying that normativity is a genuine dimension –
the point is to underline that our ethical systems are molded on the basis of an
action-theoretical framework.³

**Five Dimensions to Analyze Action**

After having clarified our subject, it is now possible to address the next
question: Which are the main dimensions that a conception aimed at
understanding and analyzing human action entails? I would like to concentrate
on five central spheres: rationality, preferences, moral dispositions, altruism,
and global orientation of action. In other words: Every ethical theory explicitly
or implicitly rests on a theory of action that specifies: (1) the type of rationality
that individuals have; (2) the kind of preferences that lead them to action; (3)
the moral dispositions that structure their behavior; (4) the kind of altruism that

---

1Cf. Birnbacher (1999: 321): ‘[…] applied ethics without social science, however rewarding as
a theoretical exercise, may easily be sterile or even harmful.’
2For a study of the different versions of the ‘naturalistic fallacy’ from Hume to Moore, see:
Frankena (1939) and Engels (1993). Searle’s discussion of the topic is not relevant for the
purpose of this paper.
3I am using ‘theory of action’ as a synonym of a ‘sociological conception of human action’ or,
for the sake of concision, ‘sociology.’ Both action theory and sociology – at least in the so-
called individualistic approaches in the social sciences – try to develop a theoretical model of
action taking into account the contribution of empirical research. Clearly, I am not considering
here other sociological approaches, like conflict theory and functionalism.
regulates their behavior toward others; and (5) the global goals that actors pursue in life and whose achievement constitutes the ‘meaning of their life’. I will briefly discuss these dimensions.¹

(1) The first dimension of analysis refers to rationality. It can either be postulated a ‘perfect’ or ‘complete’ rationality, such as it was the case in the Homo-oeconomicus model and then employed in game theory², or, on the contrary, a ‘bounded rationality,’ such as it was conceived by Herbert Simon (1955) – a conception further developed by cognitive psychology (see Festinger 1957). In the first case, the individual is construed as a strategic actor who has full information about the alternatives to choose and who can select the option that maximizes her utility. In the second, it is assumed that the actor has imperfect information and a bounded rationality: she chooses not the best option, but just a ‘satisficing’ one. Moreover, according to Festinger, the option she chooses is the one that is in consonance with the rest of her beliefs system. In any case, social actor’s ability to calculate and process the available information in real situations is seen as less than ‘perfect’ – what leads to the development of new models of rationality (for example, Vanberg’s ‘paradigm of program-based behavior’, cf. Vanberg 2002).

The first dimension of analysis entails a key sub-domain. In fact, it can be assumed that human beings possess exclusively an instrumental rationality or that they also have a type of rationality that we can call ‘practical’ (with reference to Kant).³ The inclusion – or exclusion – of the practical rationality as an action-theoretical assumption conducts to very different conceptions of normativity. In the first case, rationality is understood along the lines of David Hume’s conception as the cognitive capacity to choose the most suited means to achieve whatever goal the individual may have.⁴ In the second case, rationality is seen as the main source of normativity: it leads the individuals both to universalize the moral judgments they use and to act according to the universalized moral norms out of a ‘sense of duty.’

(2) The second dimension of analysis that I would like to refer to concerns the kind of preferences that actors may have. It can be presupposed that actors have exclusively preferences of a materialistic type, or that besides materialistic desires and needs people may also have other, more ‘idealistic’ kinds of preferences – at least to a certain extent. The classical Homo-oeconomicus model makes the assumption that human beings try to realize

¹In characterizing these five dimensions of analysis, I will use as a starting point the classical Homo-oeconomicus model and the criticisms advanced in the contemporary sociological literature aimed at correcting and expanding that initial framework. Although much criticized and disconfirmed by numerous empirical experiments, the Homo-oeconomicus model has represented a major effort in the theory of action and social sciences to scientifically conceptualize human action. See Kirchgässner (2008).
²See von Neumann & Morgenstern (1944) and Hempel (1961).
³In this sense, some authors prefer to differentiate respectively between ‘rationality’ and ‘reasonability’, and thus between a ‘rational’ and a ‘reasonable’ decision, cf. Rawls (1999: 365-372).
⁴‘Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.’ (Hume 2000: 413)
only materialistic preferences, especially the desire to increase the personal well-being and the prestige or social reputation.¹ In later models that seek to expand the framework of the Homo-oeconomicus it is assumed that human beings, at least under favorable conditions, will pursue goals of an altruistic and idealistic kind. (In his later writings, Nozick (1995) even speaks of a ‘symbolic’ as opposed to a ‘material’ utility, arguing that the search to increase the symbolic utility plays a decisive role in human behavior.) To abide by norms of fairness even when it is detrimental to the own material benefit; to pursue an increase of the well-being of others even when they cannot reciprocate with ego; or to act in conformity with the own conception of a meaningful life even if it does not imply an increase on prestige or wealth: these examples show that behavior can be motivated by preferences that transcend the narrow sphere defined by the own material well-being.²

(3) The third dimension of analysis concerns the kind of moral dispositions that can be presupposed in the individuals, for example, (universal) sympathy or fairness. Again, in the classical Homo-oeconomicus model no moral disposition altogether was supposed in the actor or it was only assumed that agents have a general disposition to reciprocate and cooperate with others if it is in one’s own long-term interest. Once more, in contemporary conceptions of human action it is generally supposed that human beings have a disposition to act fairly – independently of whether it is in the actors’ interest or not. It is nevertheless worth noticing that this moral disposition can be characterized in different ways. While some authors³ speak of a disposition to act fairly above all in face-to-face interactions, others – following Lawrence Kohlberg’s work (1963; 1969) – assume a pervading disposition to act according to principles of fairness independently of the specific kind of social situation which the actors are part of. There is also a tendency – especially among authors coming from virtue ethics like Blum (1994) – to assume that actors in general do possess and develop a wide range of caring dispositions or virtues toward others, like sympathy, generosity, and compassion. (Here, like in all other dimensions I am discussing, it is not only important to assess if social actors may have moral dispositions, but also to what extent and under which social conditions these dispositions would be fully developed and become a leading force in the regulation of one’s behavior. This point is particularly important when we consider the undeniable presence of egoistic or even antisocial dispositions as part of real-life human behavior.)

(4) The fourth dimension relates to the sort of altruism that can in general be assumed among human beings. After the seminal studies of W. D. Hamilton (1964) and R. Trivers (1971) in the field of evolutionary biology, even

¹See Hill (2002).
²See the discussion in Kirchgässner (2008). It is worth noticing that the sociological model should always specify under which conditions – and up to which extent – one type of preferences can be more ‘intense’ that the other; if it is false to maintain that human beings are exclusively concerned with materialistic desires, it is also pointless to assume that they simply act according to ideals (be it the self-realization, the achievement of universal moral goals, etc.)
³See the seminal work of Roth (1991), and the discussion in Güth & Kliemt (2010) and Guala (2009)
philosophers reluctant to idealistic considerations assume nowadays that
human beings generally behave according to a pattern defined by kin and
reciprocal altruism. Thus, the assumption of an unconditional egoism has been
abandoned to include some restricted forms of altruistic behavior.
Nevertheless, the question is if we can additionally include among our action-
theoretical assumptions other types of altruism, for example, a behavioral
disposition to act altruistically within larger groups, be the community, the own
national society, or even the whole human race – and out of emotions and
sentiments involving love, compassion, etc. (In any case, it is important to
insist that after assuming that human beings may behave altruistically toward
strangers in situations that most probably will not benefit ego, the point is to
determine concretely under which conditions they will act according to the
pattern of universal altruism.)

(5) The fifth and last dimension of analysis that I would like to address
concerns what we can call the global orientation that can be attributed to
human action in general. Again, in the classical Homo-oeconomicus model it
was supposed that human beings try to maximize utility by choosing courses of
action according to two key preferences: the own well-being and the reputation
(and in some cases the well-being and reputation of close family members and
reciprocators) (Hill 2002, Kirchgässner 2008). As a consequence, the question
of the ‘meaning of life’ was considered irrelevant or it was simply assumed that
utility maximization constitutes the pervading goal of human life. On the
contrary, in current approaches it is supposed that human beings try indeed to
satisfy their needs and wishes but as part of a more global plan of life they have
(see Wolf 2009; Rawls 1999: 347-396). According to this point of view, human
beings have essentially a conception of a good and meaningful life that
establishes for them a certain plan of life. In this way, the maximization of
utility derived from an increase of well-being or reputation is (at best)
subsidiary to the realization of their life-project, and not a goal in itself.

To summarize: I have presented five key dimensions that are central when
conceptualizing human action. In characterizing and specifying these
categories, the Homo-oeconomicus model serves as a starting point.
Nevertheless, revisions and new developments of the initial model inside the
framework defined by the individualistic approach in the social sciences lead to
a more complex but realistic picture of human action.1

As already argued, the ethicist, in developing her own normative system,
makes use of a certain conception of human action, but generally in an implicit
way. She would make, for instance, some assumptions regarding the type of
rationality that human beings possess or the kind of altruism that regulates their
behavior, but those dimensions remain mainly vague or do not receive a
systematical treatment grounded in the knowledge supplied by the social
sciences. I maintain that by addressing these sociological aspects and

1In this sense, we can consider the theory of action as an interdisciplinary field in which
philosophers as well as social scientists (belonging to the individualistic tradition) try to
develop a general model to understand and explain human action as the primary step for a later
understanding of social phenomena at a macro level.
discussing them with help of current developments from the field of action and sociological theory we can contribute decisively to normative ethics. Contemporary social theories can provide key tools for the analysis of a given conception of human action and thus have a bearing on the construction and evaluation of normative systems. To show the way it concretely works, I will briefly present a case study.

A Case Study: Contractarianism

John L. Mackie is rather well known in the meta-ethical discussion because of his ‘error theory,’ but the rest of his contribution to normative theory remains to a large extent unknown, although his three moral studies represent one of the clearest formulations of contractarianism and conventionalism – an approach inspired by Hobbes’s *Leviathan* and Hume’s *Treatise* (Mackie 1977; 1980; 1985).

According to Mackie, moral skepticism is right in pointing out that moral values are a human creation (and that are thus ‘subjective’), but skeptics make a mistake when inferring from there that values and moral norms are just personal or idiosyncratic inventions that may be changed whenever one wishes – in other words, that they lack social stability. Inventing right and wrong – to quote the subtitle of Mackie’s first moral study – has been a long, complex, and deep-rooted process of cultural evolution. As a result we have, on the one side, the moral dispositions that individuals internalize and, on the other, the moral institutions that through political and social arrangements safeguard the observance of norms.

In this way, Mackie considers normativity as (just) a ‘supervenience’ emerged out of a long process of trial and error at the end of which it has been possible to reach some agreements to regulate human behavior – what has made life and cooperation at least to some extent possible. Put differently, normativity rests on individual’s basic desires and needs, and can be traced back to conventions aimed at protecting the individuals’ personal sphere. For Mackie, there are no other sources of morals.

Once established that the dimension of normativity can be reduced to social conventions, Mackie suggests that moral norms have only a prudential character. Accordingly, the task of the moral philosopher is to develop a rational and clearly specified moral system, establishing first of all a series of basic rights that protect the individual sphere and subsequently a minimal normative system.

---

1The notion ‘subjectivism’ characterizes Mackie’s meta-ethical position. Some of his central insights regarding contractarianism and conventionalism can be found in the contemporary discussion, for example in Gauthier’s *Morals by Agreement* (1986). (I cannot discuss here the similarities and differences between the proposals of both authors.)

2For the distinction between (Mackie’s) ‘subjectivism’ and ‘plain subjectivism’ – a position that can be found in many emotivists –, see Stroup (1981).

3In a sense, the task of the ethicist is just to redesign – through rational thinking – the very system of conventions that emerged from cultural evolution through a long, blind, and sometimes (from our present point of view) inefficient process.
amount of obligations whose observance safeguards those rights. All other type of ‘broader’ moral systems – like utilitarianism, Rawls’s contractualism, and of course the moral systems of major religious movements – not only lack theoretical justification, but make also of morals a ‘burden’ that exceeds the limits that the individual’s motivational framework, such as fixed by human nature and cultural evolution, can hold.

In Mackie’s opinion, the only way to reconstruct an adequate normative theory consists in imagining an original situation populated by rational and self-interested individuals, each of them representing a threat to the others (Mackie 1985: 114; 185). These individuals are further construed as having similar basic desires and needs, like the desire to act freely and enjoy the own property. In that context, the first step is to establish a series of basic rights, first of all the right to personal integrity and liberty, and then the duties whose fulfillment would secure those rights. If an obligation cannot be traced back to a supposed arrangement between rational and egoistic individuals, it is declared to devoid of any justification and plausibility.

The result of that procedure is a minimal moral system, a liberal constitution, and a set of guidelines for the establishment of a minimal state as the institution that best secures the compliance to those basic but decisive social norms.¹

It is not possible here to present Mackie’s ethical theory with further detail. It suffices to stress that his way of thinking represents a major challenge to normative approaches like the utilitarianism of Richard Hare and Peter Singer, and to the whole school of thought inspired by Rawls’s work. Mackie’s theory can be thus considered as one of the most solid defenses of libertarianism in current political thought.

 According to the approach advanced in this paper, Mackie’s theory cannot only be discussed in philosophical terms, but – and especially – by highlighting and analyzing the conception of human action that works as the basis of his normative developments. There, the individuals are understood as rational actors in the sense that they make use of an instrumental rationality to reach their egoistic and mostly materialistic goals. Consequently, intensive altruistic or idealistic preferences can only be attributed – if any – to a minority of human beings – a too small part of the population to be considered in any serious theoretical construction. The majority of human beings behave according to the pattern of the so-called kin and reciprocal altruism. In Mackie’s implicit conception of action, human beings can indeed internalize moral dispositions, but only in so far as those dispositions secure a type of cooperation that returns in the same individual’s benefit. To demand the development of other kinds of moral dispositions seems in this view unrealistic or even hypocritical. Finally, Mackie’s conception presupposes that the global goal of human life consists in the satisfaction of the own desires and needs, and

¹Nozick’s famous book Anarchy, State and Utopia was published in 1974, just three years before Mackie’s Ethics, and can be seen as an anticipation of many of the consequences of Mackie’s position for political thought. (The need of a (minimal) state is, by the way, one of the most salient differences with Gauthier.)
in the increase of the own reputation.\textsuperscript{1}

Although Mackie was fully aware of some shortcomings of the classical Homo-oeconomicus model (Mackie 1985: 120-131; 206-219), his conception of the human being remains very close to that initial framework. As a consequence, many of the criticisms and further developments advanced to that model can be applied to Mackie. I would like to mention just three of these topics. The first one refers to Mackie’s refusal to admit any form of rationality of a practical kind. It is true that many times the assumption of a practical rationality constitutes a pretext for introducing surreptitiously a large set of duties and obligations that corresponds above all to the philosopher’s own moral or religious conception. Nevertheless it is worth considering the possibility of conceiving reason as a source of normativity. To say that human beings are reasonable beings means that they are able to universalize and follow the moral judgments they use everyday’s life – a process that leads to the establishment of moral norms that go beyond the (narrow) sphere defined by enlightened self-interests.\textsuperscript{2}

The second point concerns the dimension of moral dispositions. Here it is also the case that – as Mackie would warn – we should be careful when assuming moral dispositions among social actors. In philosophy it is not infrequent to end up having a too idealistic picture of human beings. However, I maintain that Mackie concentrates too much on how human beings de facto are – and not on how they could be, i.e., \textit{on the capacities and potentials} that human beings do possess but that are not fully developed \textit{under present social conditions}.\textsuperscript{3} Individuals – so can be argued – have much potential (of a moral and of other kinds) that are currently not fully developed or not developed at all.

The third point refers to the view that Mackie holds concerning the general goal of human life. To maintain that human beings exclusively try to increase their material well-being and social status is an assumption that surely helps to explain some aspects of human behavior, especially the one of social actors in economic interactions (for example, buying and selling in an anonymous market). It is also true that, on the other side, many philosophers have a quite heroic and idealistic picture of human beings. However, an adequate conception of human action should probably assume that individuals possess a (culturally mediated) conception of a good and meaningful life – and that they essentially seek to live according to that conception that renders their lives (at least in their eyes) ‘worth living’ (Wolf 2009).

\textsuperscript{1}Mackie’s implicit picture of human beings does not differ much from the one that evolutionary biologists and game theorists have when analyzing interactions. It is true that he pays attention to cultural evolution, but in his work history seems to be – to paraphrase Clausewitz – nothing but a continuation of natural evolution by other means.
\textsuperscript{2}Cf. Singer (1981: 147): ‘I have suggested that reason is not powerless. On the collective level, once we have begun to justify our conduct publicly, reason leads us to develop and expand our moral concerns, drawing us on toward an objective point of view.’
\textsuperscript{3}See the discussion in Blum (1994, chapter 7: ‘Virtue and Community”), and in the reflections on the ‘sense of justice’ and ‘a well-ordered society’ in Rawls (1999: 397-449).
Concluding Remarks

In this paper, I have advanced what I consider to be an original way of approaching to the discussion in contemporary normative ethics. In a nutshell, the point is to focus systematically on the sociological assumptions upon which these theories rest. My proposal is that before engaging in the construction of a normative system, the philosopher should not only settle down questions concerning meta-ethical aspects but also gain a clear and robust conception of human action. Two are the reasons: Firstly, it is possible to shed new light on the ethical discussion if the action-theoretical elements of normative theories are properly highlighted and scrutinized. I have shown how this approach works in the case of contractarianism, but a similar analysis can be applied to other normative theories, like utilitarianism and Rawls’ contractualism.\footnote{As a hypothesis, it can be argued that none of the quoted moral theories rests on an adequate conception of human beings – a provocative statement intended to further new developments in normative ethics.}

Secondly, in the last decades theory of action and sociological theory have consolidated as recognized fields of research through key theoretical and empirical developments. This enables the scholar to decide up to which point an ethical theory bases on an adequate conception of human beings.\footnote{‘In arguing that a co-operation between ethics and social sciences is called for in the field of applied ethics I do not want to imply that the more theoretical questions of normative ethics can safely be left to philosophy. On the contrary, it seems that even some of the fundamental questions of general ethics cannot be dealt with satisfactorily without reference to anthropological matters of fact which stand in need of empirical elucidation and confirmation.’ (Birnbacher 1999: 229-230, my italics)}

As a practical consequence, I would like to conclude stressing the need of a more intensive and better structured interdisciplinary co-operation between ethics and social sciences. Although interdisciplinarity is generally considered as a desirable trait of any research project, in reality both philosophers and social scientists continue to work separately. More concrete efforts in this direction would undoubtedly benefit both groups of researchers.\footnote{In this article I have tried to show how ethics can benefit from sociology, but it would be equally important to discuss what sociology can gain from a systematic cooperation with ethics. Lacking space for an appropriate treatment of this topic, I restrict myself to state the following: Although sociology intends to carry out scientific (and thus value-free) research on social phenomena, most sociologists include normative considerations in their studies. The problem is not this feature as such (in so far as it does not compromise the objectivity of the research), but that those considerations are most of the time vague and lack a proper philosophical treatment. That is exactly the point where ethical theory can be of undisputable utility to sociology.}

Literature

The Role of Sociological Assumptions in the Construction of Ethical Theories


Vol. XIX, Nr. 2 (pp. 213-234).


Every conception of happiness, philosophic or otherwise, contains within it the seeds of revolution, the promise of utopia. The only question then is whether this promise is made explicit or remains implicit, whether it is drawn out into the light of day or buried and repressed. At this point I will not offer a defense of this claim, trusting that the defense will become clear in what follows, but rather state that it is my intention to examine an article recently published by Simon Critchley in light of it.\(^1\) As every promise that is allowed to remain buried becomes an apology for the status quo,\(^2\) it is my intention to unearth the utopian promise already at work within this concept.\(^3\)

My critique will proceed in several steps. First, for the benefit of my readers not already familiar with Critchley’s article, I will give an account of his conception, hopefully doing it the justice it deserves. Next, I will elaborate why this conception seems to fit our general conception of happiness, that is, I will attempt to build a case for Critchley’s concept. Elaborating this case will necessarily require some discussion of the background against which it operates and which makes it compelling. That background is, quite naturally, the social world with which we are all more or less familiar. Third, and here is where I depart from Critchley, I will attempt to draw out the contradiction already at work between the concept of happiness elaborated by Critchley and the social world in which this concept operates. It is my contention that this contradiction is already at play, and that Critchley is not unaware of this. It is this contradiction which makes the concept compelling. But what Critchley’s article fails to do, and what I hope to do, is to reflect on the meaning of this contradiction, for it is here that we find the utopian promise of the concept.\(^4\)

\(^1\)While this examination will necessarily push us beyond the scope of Critchley’s article, I believe that it will take us in a direction that is necessary if we do not wish to surrender the concept of happiness to the forces of the status quo.

\(^2\)That is, insofar as one neglects as much as note the critique of the status quo that is already inherently at work in the concept of happiness, one suggests that no critique is necessary or available. The conclusion all too easily drawn from this omission is that ‘everything is fine the way it is’. This is the conclusion I would like to avoid.

\(^3\)If Adorno is correct in suggesting that, “All reification is forgetting”, and criticism really means the same as remembrance - that is, mobilizing in phenomena that by which they have become, and thereby recognizing the possibility that they might have become, and could therefore be, something different’, then the critique which I propose should be read as a work of remembrance. Theodore Adorno, *Introduction to Sociology*, p150.

\(^4\)As Marcuse writes, ‘To the degree to which the established society is irrational, the analysis in terms of historical rationality introduces into the concept the negative element - critique,
Finally I turn to an evaluation of Critchley’s concept of happiness. For while there are certain respects in which I find it compelling, especially when set against our familiar social environment, I find it comes up short in certain other regards. That is, not only is there a contradiction at work between Critchley’s conception of happiness and the social world in which we all live, this concept is equally inherently contradictory, pointing towards utopia at the same time as it begs its apologies from the status quo.

**Part 1: Critchley on Happiness**

Critchley begins his article by noting that ‘For the philosophers of Antiquity, notably Aristotle, it was assumed that the goal of the philosophical life — the good life, moreover — was happiness and that the latter could be defined as the bios theoretikos, the solitary life of contemplation’. While this ‘solitary life of contemplation’ might not seem to be the obvious front runner in today’s race for the ‘good life’, given how ‘[t]he rhythm of modern life is punctuated by beeps, bleeps and a generalized attention deficit disorder’, it is Critchley’s aim to make ‘the idea of happiness as an experience of contemplation’ appear less ridiculous than it might on the surface.¹

Curiously, Critchley does not begin to make his case by examining the role of contemplation in Aristotle’s work on ethics, nor does he attempt to draw out the connection between contemplation and happiness at play in Aristotle’s work as one might be forgiven for expecting. Instead, Critchley jumps ahead a few years to introduce a passage from Rousseau’s *Reveries of a Solitary Walker*, a passage that is, in Critchley’s words, ‘as close to a description of happiness as I can imagine’.² Forgive me then as I, following Critchley, quote the passage in full.

> If there is a state where the soul can find a resting-place secure enough to establish itself and concentrate its entire being there, with no need to remember the past or reach into the future, where time is nothing to it, where the present runs on indefinitely but this duration goes unnoticed, with no sign of the passing of time, and no other feeling of deprivation or enjoyment, pleasure or pain, desire or fear than the simple feeling of existence, a feeling that fills our soul entirely, as long as this state lasts, we can call ourselves happy, not with a poor, incomplete and relative happiness such as we find in contradiction, and transcendence’. Herbert Marcuse. ‘The Catastrophe of Liberation’, p103. It is the last three that I hope to draw out of Critchley’s conception of happiness.

¹Crichtley, ‘Happy Like God’.

²We may account for this ‘leap’ by suggesting that Critchley may have correctly wagered that Rousseau would be much closer to us than Aristotle, and hence a better partner in his difficult task of making a case for contemplation as the highest form of the good life available to us. While Critchley may have been correct in his wager, it is my contention that Rousseauian reverie is not an adequate substitution for Aristotelian contemplation and the latter gives us good grounds on which to critique the former.

322
the pleasures of life, but with a sufficient, complete and perfect happiness which leaves no emptiness to be filled in the soul. (Emphases Critchley)¹

Although reverie is a distinct experience, perhaps even one with which we are all familiar, capturing it in words is not quite so easy. Critchley does an admirable job, characterizing this state of reverie as follows, ‘one is awake, but half asleep, thinking, but not in an instrumental, calculative or ordered way, simply letting the thoughts happen, as they will’.² Noting that his experiences of reverie come not while floating, like Rousseau, in a little boat on a lake, but rather while sitting by the sea, listening to ‘the never-ending drone of the surf’, Critchley adds, ‘At moments like this, one can sink into deep reverie, a motionlessness that is not sleep, but where one is somehow held by the sound of the surf, lulled by the tidal movement’.³

What then makes this experience, this state of reverie, the epitome of happiness? While I do not pretend to give an exhaustive account, three features stand out. Reverie, and by extension happiness, is non-teleological, distraction free, and finally, and this is the point that Critchley emphasizes, marked by the experience of ‘timelessness’, of being ‘in time but without concern for time’. This last point is perhaps the most important, and Critchley is right to emphasize it, for it is this one which seems to include the other two, and, would, I suspect, include any other features of happiness which we might be able to draw out of this moment of reverie.

Critchley describes Rousseau in his boat as ‘floating’ and ‘drifting’, and himself by the sea not so much as thinking, but rather as letting ‘the thoughts happen, as they will’. Insofar as we can imagine either ‘drifting’ or ‘floating’ as modes of travel, they are modes of travel completely detached from any thought of either the destination or the means of getting there. This is analogous to Critchley’s ‘thoughtless thinking’, where he seems to more the passive observer of ideas floating through his head, than an active thinker engaging with those thoughts. If reverie can be said to have a logic at all, it would be the logic of dreams, and not the instrumental logic of our everyday, waking lives. The ‘thoughtless thinking’ of reverie moves in those moments of ‘waking sleep’. This ‘waking sleep’ is also a time free of distraction, or as Critchley quotes Rousseau, a moment in which ‘the soul can find a resting-place secure enough to establish itself and concentrate its entire being there’.⁴ This is a picture of one’s ‘collecting one’s self’ or of returning to itself after having been diffused through the world. The soul is at rest not simply with itself, but in itself, free from its entanglements with the outside world.

Rousseau connects the experience of timelessness with the ‘simple feeling of existence’, which Critchley understands as a ‘sentiment of momentary self-

---

¹Rousseau, Jean-Jacques, Reveries of a Solitary Walker, quoted from Critchley, ‘Happy Like God’.
²Critchley, ‘Happy Like God’.
³Ibid.
⁴Critchley, ‘Happy Like God’.
sufficiency’. We are ‘in time’ but not ‘concerned with time’, but to not be concerned with time is to not see ourselves as ‘in time’. Moments of reverie are those which break the spell of time, and breaking the spell of time is perhaps the secret to living fully embodied in the present ‘with no need to remember the past or reach into the future’. This ‘simple feeling of existence’ is really not quite so simple at all, for what we experience in these moments when the spell of time is broken is not simply the eternal present known as eternity, but the ‘feeling of existence’, of what it means ‘to be’.

**Part II: Backdrop, or ‘The World’**

Just as man does not take his bearings in reference to an abstract conception of space, but always in reference to some particular space, so to our conceptual life evolves against the ever changing backdrop of our material life. We are not simply happy, but happy at some place, in some time, in some situation. This is tacitly acknowledged by the series title in which Critchley’s article appeared, ‘The Pursuit of What Matters in Troubled Times’. Having given an account of Critchley’s take on reverie and happiness, we now turn to a consideration of the backdrop against which this conception is elaborated.

Luckily, we do not have to look to far afield to locate this backdrop, the modern social world in which we live and against which the concept of happiness is elaborated. Critchley himself succinctly captures what he takes to be the essence of modern life. ‘Our lives are filled with the endless distractions of cell phones, car alarms, commuter woes and the traffic in Bangalore. The rhythm of modern life is punctuated by beeps, bleeps and a generalized attention deficit disorder’. This is the social world ruthlessly, and brilliantly, satirized in *The Gods Must be Crazy*, wherein the technological gadgets which were supposed to make our lives more convenient, indeed, which were supposed to make our lives more free, seem to possess a life of their own, exerting more control over us than we over them. It is a world that relentlessly forces itself upon us, a world where our ‘private time’ is increasingly at the mercy of a world that does not seem to want to release us from its grasp. The laptop that makes it possible for me to work anywhere chains me ever tighter to the machine as every possibility is transformed, as if by alchemy, into an expectation. While we slept a neat reversal has occurred, as the Kantian dictum of ‘should implies can’ now reads ‘can implies should’. This is the world of total ‘social rationalization’, a world which is ‘rational in its means but not in its ends’.

---

1 Lamenting the impact of the social world on our ability to live fully embodied in the present, Vaneigem writes, ‘most of the time, alas, the whirligig of memory and anticipation inhibits both the expectation and the experience of the present by sweeping it along in the millrace of dead time, a sequence of hollow instants’. Raoul Vaneigem, *The Revolution of Everyday Life*, p230.

Of course, it’s not simply the world’s forcing itself upon us, ‘beeping and bleeping’, that is to blame for the distractions and disruptions in our lives. For if it is true as Critchley’s states that modern life is best characterized as suffering from ‘a generalized attention deficit disorder’, it is equally true that we seek these distractions out and create them for ourselves. As anyone who has spent time with a ‘modern teen’ may be able to attest, it is not simply the case that they are easily distracted, but rather more pervasively, they live lives that are completely decentered. Speaking on the phone, updating one’s social networking site, viewing a movie - these are not things that happen sequentially, but simultaneously, often in the company of others. To really get a picture of the social world created by technology, and the ‘distractions’ it promises, we do well not to look simply amongst ourselves, at people of our generation, but at the next. For therein lies the face of the ‘modern world’.

At the same time we might wonder whether this depiction of our ‘modern world’ is entirely sufficient. For not only do we live amidst constant distraction and disruption, the ‘beeping and bleeping’, we also live amidst a generalized anxiety that not infrequently erupts into outright panic. The ever-impending threat of environmental collapse, terrorist attacks, economic meltdown, flu pandemics (just to name the top few) color our social world just as surely and pervasively as do ‘the endless distractions of cell phones, car alarms, commuter woes and the traffic in Bangalore’. It is also a world in which your chances of survival still depend to a large degree on where you happened to be born. Comparatively, the traffic in Bangalore looks positively benign.¹

Part III: Contradiction

Given the descriptions given above of happiness, i.e., the experience of reverie, and the social world in which it operates, it is not surprising that the two should be in conflict. My goal here then is not to create conflict where none exists, but as Hegel might have put it, to merely watch as this conflict takes shape. In claiming that happiness and the existing social world are in contradiction to one another, I do not take myself to be making a claim that Critchley, or Rousseau for that matter, would not likewise endorse. I do not think that I claim to much either if I state that both would equally agree that it is this contradiction that accounts, at least in part, for our equating reverie and happiness.

We get a sense of the contradiction at work by focusing on the space, real or metaphorical, wherein reverie occurs. For while it is true that reverie occurs

¹Critchley himself provides a more gripping depiction our current state of affairs writing of political disappointment in his work Infinitely Demanding Ethics of Commitment, Politics of Resistance. ‘In the latter [political disappointment], the sense of something lacking or failing arises from the realization that we inhabit a violently unjust world, a world defined by the horror of war, a world where, as Dostoevsky says, blood is being split in the merriest way, as if it were champagne. Such an experience of disappointment is acutely tangible at the present time, with the corrosion of established political structures and an unending war on terror where the moods of Western populations are controlled through a politics of fear managed by the constant threat of external attack’. Simon Critchley, Infinitely Demanding Ethics of Commitment, Politics of Resistance, p3.
within the existing social world, there is a very real sense in which it occurs in a space equally removed from it. This also has the meaning that in the experience of reverie, one transcends the given social reality while remaining rooted in it. Critchley’s examples are highly illustrative in this regard. Rousseau rowing away from the confines of city life to seek refuge on the lake, Critchley himself, like a modern day Ishmael, pulled by the promise of the sea, Wordsworth wandering ‘lonely as a cloud’ - each of these metaphorically express the conceptual space of reverie as existing just outside, or rather, as an isolated pocket within, the given social world. We could describe reverie as an internal transcendence of the social world, but as transcendence, it is fragile and temporary at best, and one that remains connected to that social world, just as Rousseau’s wilderness eventually must border on the city. And just as city and wilderness are a conceptual pair which are only meaningful together, so too does the force of reverie draw its strength from its revolt against the social environment. We have characterized reverie as a temporary, internal transcendence. We could equally call it ‘flight’ or ‘escape’.

As escape however, reverie is doomed to fail, for in its contradiction with the social world, it is the latter which is the more dominant partner. Critchley captures this with pithy eloquence, ‘The cell phone rings, the e-mail beeps and one is sucked back into the world’s relentless hum and our accompanying anxiety’. The social world intrudes upon our moment of reverie and reverie, which leads a fragile existence, cannot but succumb to the intrusion. Against the demands of the social world, this moment must collapse; escape must end in return, however unwilling, however unwelcome.¹

Part IV: Happiness and Promise

‘Today people no longer blame their sufferings on the hostility of nature, but on the tyranny of a perfectly inadequate and perfectly anachronistic form of society’.² While this is true, at least for most of us in the ‘developed world’ who no longer feel threatened by floods³ but by the twin evils of work and joblessness,⁴ Critchley’s article not only fails to take this properly into account,

¹On Marcuse’s reading, these fragile moments of reverie are allowed to exist only insofar as they serve the very social mechanism from which they flee. He writes, ‘For the immediate producer the restriction of enjoyment operates immediately, without any moral mediation, through the working day, which leaves only a brief period of ‘leisure time’ for enjoyment and puts it in the service of relaxation and the recreation of energy or labor power’. Herbert Marcuse, ‘On Hedonism’, p167.
²Vaneigem, The Revolution of Everyday Life, p77-78.
³Even with the flooding caused by hurricane Katrina in New Orleans, I would wager that most of those affected by the storm would cite the poor response of the government, and not the storm itself, as the immediate cause of their sufferings.
⁴A few lines from The Smiths captures this dilemma well. ‘I was happy in the haze of a drunken hour/but heaven knows I’m miserable now/I was looking for a job and then I found a job/and heaven knows I’m miserable now/In my life/why do I give valuable time/to people who don’t care if I live or die’. For far too many who have had their passions and imaginations
but more problematically, obfuscates the distinction between the ‘social world’ and the ‘natural world’. Given this failing, reverie, as the shape of happiness, fails to offer a critique of the social world and instead becomes an apology for it.\(^1\)

We return again to Critchley’s account of modern life. He writes, ‘Our lives are filled with the endless distractions of cell phones, car alarms, commuter woes and the traffic in Bangalore. The rhythm of modern life is punctuated by beeps, bleeps and a generalized attention deficit disorder’.\(^3\) I have already indicated that while Critchley’s depiction of modern life is true as far as it goes, it does not do justice to the pervasive injustices which are equally, and more damning, features of the modern world. But there is something else at work here, and that is Critchley’s acceptance of the ‘givenness’ of these features of the modern world. They are singled out as naturally occurring artifacts, no different from rocks, trees, draughts, or floods. They are just what we happen to ‘find’ in the modern world. There is no indication that these features of modern life, these ‘beeps and bleeps’ which we just happen to find, do not occur naturally in ‘the world’ as do rocks and trees, but rather occur through our organization of the social world. That is, Critchley gives no indication that this social world as we currently find it has only the shape which we have given it, which means conversely that we could have given it a different shape and that the shape which it will take in the future will be the shape which we have decided to give it.

We see Critchley’s equivocation at work again as he ends his article. ‘And then it is over. Time passes, the reverie ends and the feeling for existence fades. The cell phone rings, the e-mail beeps and one is sucked back into the world’s relentless hum and our accompanying anxiety’.\(^4\) If Critchley extended the ‘givenness’ of the natural world to cover the social world as indicated above, in this brief passage he erodes the distinction completely. One is ‘sucked back into the world’s relentless hum’ [my emphasis], and here it might just as well be the natural world as the social world. As a description of ‘the world’, ‘relentlessly humming’ functions well to cover the distinction between the natural and the social, for in their own ways, both ‘hum’. Sitting in my

---

1I recognize that this distinction between the ‘social world’ and the ‘natural world’ may sound a bit hollow, especially as more and more of ‘nature’ gets taken up into the social. True as that may be, I believe the distinction is still necessary insofar as ‘natural’ always indicates that which is on the other side of human dominion and presents itself as simply ‘given’, or as brute fact, even if it turns out that in the future which is rapidly approaching nothing is any longer on the other side.

2As Marcuse writes, ‘All joy and all happiness derive from the ability to transcend Nature - a transcendence in which the mastery of Nature itself is subordinated to liberation and pacification of existence. All tranquility, all delight is the result of conscious mediation, of autonomy and contradiction. Glorification of the natural is part of the ideology which protects an unnatural society in its struggle against liberation’. Herbert Marcuse, *One Dimensional Man*, p237-238.

3Critchley, ‘Happy Like God’.

4Ibid.
kitchen in my old apartment in Brooklyn, listening to the relentless humming
of the traffic on the BQE, I often imagined that it was the humming of the
waves crashing along the shore far off in Coney Island.

By collapsing the distinction between the natural world, the world that truly
is ‘given’ and that confronts us as a fact, and the social world, the world made
and organized by the conscious and unconscious impulses of real people,
Critchley invites us to see the social world not as the work of our own hands,
and not as something which we have the ability to alter. The social world may
change, as does the natural world, but as the distinction between the two
collapses, we are lulled into thinking that these are changes which take place
above and beyond us, and of which we can have no control. We may escape
into momentary reverie and find our happiness there, but this too will pass and
soon enough we will find ourselves ‘sucked back into the world’. Of course,
there is another option, one that Critchley fails to mention, and that is to use
these moments of temporary reverie, these moments of happiness, to critique
the existing social world. For the truth, which gets buried in Critchley’s article
is that the social world, the world we live in, takes on the appearance of
operating above and beyond us only insofar as we treat it as simply given, as
out of our hands. The social world is not simply some ghost, pulling the strings
behind our backs, but is equally that world that we create and recreate as we
live our lives in common.

If happiness today takes the shape of reverie as Critchley maintains, it does
so insofar as it takes its stand in opposition to the social world as it exists
today. Submerged in this opposition, and doubly so in Critchley’s article, is a
demand for a different arrangement of the social world, a promise of a better
world lying on the other side of this one. Although it would be premature to
attempt to describe these new arrangements, this promised ‘better’ world, the
shape of happiness as we find it today contains some pointers. For what we

---

1 Just like Tom Cruise and Nicole Kidman in Stanley Kubrick’s *Eyes Wide Shut* who, after
taking flight into the dangerous worlds of desire and imagination, a flight which threatens to
break up the comfort of their everyday existence, resolve to shoehorn these desires into that
very everyday existence from which their escape had derived such passionate intensity. In both
cases a reified social order is seen to triumph over the demand and promise of happiness.

2 In case I have presented this point too one-sidedly, there is a very real sense in which the
social world that we inherit and inhabit does function like a ‘ghost pulling strings behind our
backs’. But while we can admit this, we do not need to give up on the second claim, that is,
that the social world is equally, and in just as real a sense, ‘the world that we create and
recreate as we live our lives in common’. To neglect the second half of this claim is to
surrender to the first half. Critchley’s article invites us to make just this surrender.

3 Karl Marx’s remarks on emancipation and reform are appropriate here. ‘The internal
difficulties seem almost greater than the external obstacles. Even if there is no doubt about the
‘whence’, there is only all the more confusion about the ‘whither’. Not only has general
anarchy broken out among reformers, each will have to admit to himself that he has no exact
idea of what is to be. But this is precisely the advantage of the new direction, namely that we
do not anticipate the world dogmatically, but seek to find a new world through criticism of the
old... If the building of the future and its consummation for all time is not to be our work, then
it is still clearer what we have to accomplish at this time - I mean the relentless criticism of all
existing conditions, relentless in the sense that the criticism must not be afraid of its results and
just as little afraid of conflict with the powers that be’. Karl Marx, ‘From A Correspondence of
Thoughts by the Sea, Reflections on Happiness

find in reverie is a plea for a space free from the demands of production and consumption which have so fashioned our cities, truly autonomous time, not the orchestrated and administered form which goes by the name of ‘free time’, and for reason unshackled from the purpose driven logic of the business cycle and wed instead to its truthful partner, imagination. That these demands, simple and essential as they are, read as utopian promises, where ‘utopian’ meets with derisive laughter as opposed to passionate wonder, is an indictment not just of the ‘beeping and bleeping’ world we have created for ourselves, but equally of the hold that world has on us, body and mind.

Although Critchley does not discuss this aspect of reverie, it, as the shape that happiness takes today, appears to have in its favor that it is radically democratic. Reverie, seemingly free from any basis in the material world from which it would need to draw support, is accessible not just to the Rousseaus of the world, but to the Walter Mittys as well. The CEO of any large corporation, by grace of his largesse, has no more access to these moments of reverie than does the child laborer producing goods that she will never afford in one of his sweatshops. As the disparity between the richest and the poorest among us continues to widen, we can take heart that reverie, that happiness, is available to everyone on an equal basis. Indeed, we could even imagine that for Plato’s prisoners, chained up in the cave from birth to death, alive with the illusions that they regard as truth, with shackles cutting into their flesh, the experience of reverie remained an open possibility. We can imagine a prisoner, perhaps seduced by the rhythmic dancing of the shadows against the cave wall drifting into a state similar to that which Rousseau and Critchley have described, lulled by the dancing shadows, not quite awake but not quite sleeping, thinking, but not in any calculated way, letting the thoughts ‘happen’. We can imagine that in this moment the prisoner is able to forget the chains and the material circumstances of her existence and is able, in just that moment and for just as long as the moment lasts, to experience the simplicity of her existence, in time but without concern for time, at this very time, the moment of the present which lasts without past or future.

But could we call the prisoner happy? Should we say that the prisoner has had a ‘happy life’, a ‘good life’? Certainly we might say that in this moment of

1843’, p93. In failing to carry through this ‘relentless criticism’, Critchley presents a reified picture of ‘all existing conditions’, offering not a promise of utopia but an apology for the status quo.


2As Marcuse writes, ‘At this stage, the question is no longer: how can the individual satisfy his own needs without hurting others, but rather: how can he satisfy his needs without hurting himself, without reproducing, through his aspirations and satisfactions, his dependence on an exploitative apparatus which, in satisfying his needs, perpetuates his servitude?’ Herbert Marcuse, An Essay on Liberation, p4. Of course, we cannot begin to answer this question unless we use our philosophic concepts to critique the social world in and against which they take shape. Adorno raises a similar concern, ‘the fact that people pursue their own individual interests makes them at the same time the exponents and executors of that same historical objectivity that is ready to turn against their interests at any moment and thus may assert itself over their heads’. Theodore Adorno, History and Freedom, p27.
reverie, for as long as it lasted, that the prisoner was happy. Certainly. Or should we? Do we not threaten to drain the meaning of happiness by applying it in such fashion? Do we not threaten to legitimize the oppression of the prisoner’s situation by calling such flight happiness? The radically democratic nature of the concept of happiness stands in stark contrast to the world we inhabit today, a world wherein one ‘extra-legal’ detention center is closed and replaced by ever more ‘extra-legal’ definitions of the prisoners themselves, definitions which characterize some prisoners as ‘too dangerous to be tried’ and yet ‘too dangerous to be released’. A failure to reflect on the contradiction between the concept and the world is the mark of ideology. If happiness is possible in this world, indeed, if it is open to everyone equally, what need do we have of critique? But if the contradiction is insisted upon, another direction is possible, and that is the direction which demands that the world be made compatible with the concept, which demands democracy not for the few and privileged, but for all equally.

Part V: Counter-movements

So far I have attempted to liberate the utopian moment already at work in Critchley’s conception of happiness by sharpening the contradiction between it and the current arrangement of our social world. It has not been my intention to claim that reverie is a ‘false’ form of happiness, nor to pit one subjective experience of happiness against another. Instead, I have attempted to show both how current social conditions give rise to this particular shape of happiness and how this particular shape of happiness can help point beyond these current conditions. But this is not the only movement at work in Critchley’s article, and in the space that remains I would like to draw out two of these movements for of themselves they offer a different conception of

---

1It’s instructive to note that neither Plato nor Aristotle would be content to call such a condition a condition of happiness, even if we allowed that this moment of reverie was open to the prisoners. A happy life entails just that, a life, and happiness, at least for Plato and Aristotle could not be reduced to such a fleeting moment.

2As Marcuse expresses it, ‘That man is a rational being, that this being requires freedom, and that happiness is his highest good are all universal propositions whose progressive impetus derives precisely from their universality. Universality gives them an almost revolutionary character, for they claim that all, and not merely this or that particular person, should be rational, free, and happy’. Herbert Marcuse, ‘Philosophy and Critical Theory’, p15. These propositions will continue to have their ‘almost revolutionary character’ until the opposition between the universal claims that they make and the social world in which they operate is overcome.

3Here I diverge from Marcuse who writes, ‘It appears that individuals raised to be integrated into the antagonistic labor process cannot be judges of their own happiness. They have been prevented from knowing their true interest. Thus it is possible for them to designate their condition as happy and, without external compulsion, embrace the system that oppresses them’. Herbert Marcuse, ‘On Hedonism’, p169. I would suggest that while individuals may be adequate judges of their subjective experiences of happiness, this doe not imply that they are always adequate judges of the meaning of that happiness.
happiness than the one publicly endorsed by Critchley and so further our critique of both the shape of happiness currently on view, and the social world against which it draws it strength.

The first ‘counter-movement’ can be found in the opening lines of Critchley’s piece. He writes, ‘For the philosophers of Antiquity, notably Aristotle, it was assumed that the goal of the philosophical life — the good life, moreover — was happiness and that the latter could be defined as the \textit{bios theoreтикос}, the solitary life of contemplation’.\footnote{Critchley, ‘Happy Like God’.

2Interestingly, in another context Critchley will refer to this attempted flight from the world as a form not of happiness, but of passive nihilism. He writes, ‘Rather than acting in the world and trying to transform it, the passive nihilist simply focuses on himself and his particular pleasures and projects for perfecting himself, whether through discovering the inner child, manipulating pyramids, writing pessimistic-sounding literary essays, taking up yoga, birdwatching or botany, as was the case with the aged Rousseau. In the face of the increasing brutality of reality, the passive nihilist tries to achieve a mystical stillness, calm contemplation: ‘European Buddhism’. In a world that is all too rapidly blowing itself to pieces, the passive nihilist closes his eyes and makes himself into an island’. Simon Critchley, \textit{Infinitely Demanding Ethics of Commitment, Politics of Resistance}, p4-5. This is not to deny the passive nihilist his or her form of happiness, but rather to question whether this is the form of happiness we ought to celebrate as Critchley does elsewhere.} While I do not intend a full-scale examination of Aristotle’s conception of ‘contemplation’, I simply point out (and trust that scholars of Aristotle will bear me out) that contemplation is a markedly different type of activity than reverie. As Critchley describes, reverie is a passive form of thinking, a ‘thoughtless thinking’ wherein one simply lets the thoughts come, observing them as one might observe colors changing on a screen. Contemplation, by contrast, demands an active engagement with the material to be contemplated. Against reverie, we might call contemplation a ‘thoughtful thinking’, that is, a thinking full of thought. If the life of contemplation really were the good life as Aristotle suggests, the good life would appear to demand an active engagement with the world, and not escape from it.\footnote{Ibid.} Happiness must find itself in the world, which means that world must be made such as to allow for happiness.

Considering whether one can only be happy in isolation, Critchley writes, ‘I think that one can also experience this feeling of existence in the experience of love, in being intimate with one’s lover, feeling the world close around one and time slips away in its passing’.\footnote{Ibid.} Although Critchley chooses to focus here on the feeling of timelessness, there are other elements that equally account for the feelings of happiness one experiences with one’s lover. For what seems to lend to these intimate encounters their feeling of timelessness is not the passivity and withdraw which had characterized reverie, but rather the opposite, active engagement. Not only do these intimate encounters allow for the free play of desire and its consummation, but also for creativity, spontaneity, and autonomy. If time slips away while one is engaged with one’s lover it is not because the lovers have withdrawn from the world, but rather because in their complete immersion in their desires and each other, they have found their way back to the world. Not the world as it currently exists, but the
world as it can only be promised by happiness and demanded by love. This is a world that returns to each individual what has been stolen from her, her creativity, spontaneity, and autonomy. In a world that has been made safe for the free play of individual creativity, spontaneity, and autonomy, we can hope that happiness will come to look less like the withdrawn flight of reverie, and more like the active engagement of love.

**Works Cited**

**Literature**


**Other Sources**


Intergenerational Rights: A Philosophical Examination

Makoto Usami

In the past four decades, moral issues concerning relationships between present and future generations, particularly in the context of environmental problems, have attracted interest among moral, legal, and political philosophers. Numerous studies have been devoted to searching for the moral foundations of concern for posterity. One of the most widely accepted theories of intergenerational justice invokes the idea of the rights of future people. Inspired by Joel Feinberg’s seminal essay (1974), many authors maintain that our descendants have the right to require us to respect their interests. James Sterba (1981) argues that future generations have the right to receive the goods and resources necessary for their basic needs. Ernest Partridge (1990) defends the rights theory against several forms of early criticism. Joerg Tremmel (2006) and his colleagues explore the principles of intergenerational justice and institutional arrangements for them, presupposing posterity’s rights.

Another issue to which the notion of intergenerational rights allegedly has relevance is that of financial reparations for historical injustices, including conquest, colonization, and enslavement. In the United States, for instance, some civil rights lawyers and black activists, notably Randall Robinson (2001), propose that the descendants of slaves should be rewarded monetary compensations for their ancestors’ hardships. Demands for redress for past atrocities are also made by minorities and indigenous peoples in some other countries. In an international setting, it is increasingly argued that the current state of extreme and extensive poverty in the developing world, especially in sub-Saharan Africa, has been significantly shaped by the periods of slavery and colonialism. The same historical wrongs played a crucial role, the argument goes, in bringing about both contemporary affluence in some parts of the world and poverty in others. Accordingly, a growing number of authors (e.g., Howard-Hassmann, 2003; Butt, 2009: 97-139) defend the demand for compensations for past wrongs. Many proponents of domestic and international rectification seem to assume that the current descendants of those victimized by historical abuses have the right to compensation.

However, these claims of intergenerational rights encounter theoretical difficulties. One of the most formidable difficulties is the non-identity problem addressed by Derek Parfit (1984: 351-379; cf., Kavka, 1982). The non-identity problem refers to the fact that each person’s biological identity is contingent on the actions of that person’s natural parents, the circumstances of conception, and countless events prior to the conception. It stems from this problem that the
size and composition of a future generation in part depend on how the present generation chooses to act. It also follows that the number and identity of currently existing people are dependent on the actions of their predecessors. These implications threaten both the idea of posterity’s rights and that of current reparatory entitlements.

To circumvent the non-identity objection in a relationship between present and future generations, some rights theorists replace future individual rights with such collective rights. If the collective-rights argument is sound, all of our successors legitimately demand paramount concern of us. Others advance the threshold interpretation of harm in discussing intergenerational relationships in general. If they succeed in defending the claim for reparations for historical wrongs, citizens in some European countries will owe huge amounts of damages to those living in sub-Saharan Africa. Despite the possibly significant implications of the collective understanding of rights and the threshold reading of harm, few attempts have been made to scrutinize the solidity of these views.

To plug such a gap in the literature, this paper examines to what extent the collective conception of intergenerational rights is pertinent. I also explore the virtues and drawbacks of the threshold notion of harm. The paper concludes by suggesting that the motivation behind theories of intergenerational rights, including the two revisionist versions of them, suffers from the ambiguity of a traditional dichotomy between perfect and imperfect duties.

Some Objections to Posterity’s Rights

The idea of the rights of future generations, while attracting wide support, faces theoretical objections from various angles, although some of the objections are ill-founded. There are two opposing views on the nature of rights in the area of legal philosophy (e.g., Kramer, Simmonds, & Steiner, 1998). One is the will theory, according to which a legal right is the power to enforce or not to enforce the correlative duty as well as the power to waive the duty. The other camp advocates the interest theory, a naïve version of which equates a right with an interest intentionally protected by law. Hillel Steiner (1983: 154-155), a proponent of the will theory, insists that future persons cannot hold rights because they are incapable of exercising those rights. This criticism is ungrounded. The will theory cannot assign rights to those currently existing individuals who are incapable of meaningful choice, such as infants, senile people, and comatose persons. Nor can it satisfactorily explain inalienable rights, which are of great importance in the system of legal rights. While a naïve version of the interest theory encounters jurisprudential difficulties, a refined version solves them (Usami, 2008b: 301).

One of the well-established arguments against the theory of posterity’s rights is provided by Richard De George (1981: 159-160). He points out that future generations are not the present bearers of anything, including rights, and that they are correctly said to have a right only to what is available when they come into existence.
In responding to De George’s criticism, Lukas Meyer (2003: 145) argues that the rights that our descendants will have in the future suffice to be the grounds for our current obligations to them. His reasoning can be formulated as follows.

1. Future generations will hold rights after they come into existence.
2. The rights they will hold will be determined by the interests they will have at that time.
3. The actions and policies made by the present generation can affect the potential interests of future generations.
4. If the present generation frustrates their potential interests, this generation can be said to violate the rights they will hold.

I find Meyer’s reasoning misguided. Proposition (3) is true, as a great many observations in scientific research indicate today. Proposition (1) is also admittedly true, though one cannot identify, at present, what rights they will have. The problem with this reasoning is that proposition (2) is not false but misleading. The rights of future persons will be determined not merely by their interests but also by their natural and cultural environments, which will partly depend on the actions and policies of currently living people. If we cause the extinction of wild polar bears by continuing massive discharge of greenhouse gases, our successors will have no right to see wild polar bears, just as we have no right to see Giant Moa. Because the rights of future individuals will be in part defined by our choice of policies and actions, proposition (4) is false. Meyer’s attempted response fails, and so the point De George makes is still effective.

Non-identity Problem

One of the most powerful objections to the notion of intergenerational rights, both in the context of concern for posterity and that of historical justice, is an argument founded on the non-identity problem. Let $P_C$ represent one of the persons who will exist at a certain future point if a policy of environmental conservation is adopted at present and $P_D$ be one of the individuals living at that time if a policy of environmental destruction is now chosen. Suppose that the welfare level of $P_D$ is considerably lower than that of $P_C$. If the destruction policy is implemented, is it correct to say that the present generation is violating the rights of $P_D$?

The non-identity problem implies that $P_D$ is a different person from $P_C$, and that the existence of $P_C$ and that of $P_D$ are incompatible. The nonexistence of $P_D$ in the case of the conservation policy can be interpreted in two ways. First, a world in which $P_D$ does not exist is as good to her as a world in which she exists with zero welfare. The result is that $P_D$ is not harmed but benefited by the destruction policy, unless $P_D$ actually exists with negative welfare. It seems unsound, however, to equate the subjective value of the nonexistence of $P_D$
with her zero welfare (cf., Heyd, 1992: 37, 113) because there is no subject who enjoys \( P_D \)'s welfare when she does not exist. This leads us to the second interpretation of nonexistence: it is logically impossible to compare, in terms of welfare, one world in which \( P_D \) exists with another world in which \( P_D \) does not. Therefore, considering either of the two interpretations, one cannot help but conclude that \( P_D \) has no right to require the present generation to choose the conservation policy.

The non-identity problem also applies to the issue of financial reparations for historical injustices in both domestic and international settings. Consider the following example in the context of domestic compensation proposals. John, the African-American living in Detroit, is destitute, but his life is worth living in that he lives with positive welfare. Does John have the moral right to monetary reparations for the coerced and uncompensated labor his distant ancestors performed for many years? The answer depends on whether it is possible for him to exist in a counterfactual world in which his ancestors were not victimized by slavery. Suppose that one of John’s male ancestors, \( M_N \), was forcibly brought as a slave from Nigeria, and his wife, \( W_S \), originally came from Senegambia. Then, it is impossible for John to exist in a counterfactual world where the Atlantic slave trade did not occur and where \( M_N \) did not meet \( W_S \) at their master’s mansion in Virginia.\(^1\) This implies that should John have the right to reparation, his alleged loss would be paradoxically the very existence of him. Therefore, the non-identity problem undermines the notion of John’s right to compensations for slavery.

Let us now turn to the issue of international reparation. The Beninese peasant, Mathieu, whose ancestors were victimized by the transatlantic slave trade, lives below the World Bank’s poverty line of US$1.25 per day, but his life is worth living in terms of welfare. Does Mathieu have the moral right to require the Portuguese and the French to compensate him for the slave trade? Suppose that in the actual world with slave trading, 300 years ago, a Beninese woman, \( W_B \), married a man, \( M_B^1 \), who had escaped captivity during a war with a neighboring powerful tribe. Suppose further that in a possible world without slave trading, \( W_B \) would have preferred to marry another man, \( M_B^2 \) (or \( W_B \)'s father would have preferred to have her marry \( M_B^2 \)), but in the real world she was unable to do so because \( M_B^2 \) was held captive at the war and then taken away by slave traders. Consequently, the actual daughter, \( D_B^1 \), who was born to \( W_B \) and \( M_B^1 \), is a different person from a possible daughter, \( D_B^2 \), who would have been born to \( W_B \) and \( M_B^2 \). If Mathieu is a descendant of \( D_B^1 \), he does not exist in the possible world where slave trading did not occur and where \( W_B \) got married to \( M_B^2 \) instead of to \( M_B^1 \). The non-identity objection strikes down the

\(^1\)To circumvent the non-identity problem, it is proposed to compare the actual world with one counterfactual world in which the past interactions between victims and offenders were just and consensual, instead of with another possible world in which the acts of injustice simply did not occur (Butt, 2009: 107-115). This proposal does not help. Imagine that all of the Africans who crossed the Atlantic became well-paid workers in the two American continents. In such a possible world, the working place of each African worker depended on transactions in a competitive labor market. It is thus highly improbable that \( M_N \) and \( W_S \) met at that mansion and got married.
Intergenerational Rights: A Philosophical Examination

idea of compensatory rights.¹

Collective Rights

In the context of our relationship with our descendants, some rights theorists attempt to circumvent objections based on the non-identity problem by regarding the rights of future people as collective or group rights, not as individual ones. In her response to the non-identity objection, Edith Brown Weiss (1990: 203) says that intergenerational rights may be considered as group rights in the sense that each generation holds these rights as a group in its relationships with other generations. Such rights exist, she asserts, regardless of the number and identity of individuals making up that generation.

The idea of future collective rights appears immune from the non-identity objection, but this outlook is incorrect. Let $G_C$ denote a group of persons who will exist at a certain future point if the conservation policy is adopted at present, and $G_D$ be a group of individuals living at that point if the destruction policy is now chosen. Given the non-identity problem, it is possible—and indeed very likely in cases of remote non-overlapping successors—that all members of $G_C$ are different from any member of $G_D$. In such a case, it is quite correct to say that $G_C$ is an entirely different group from $G_D$. The non-identity problem can undermine not only the notion of future individual rights but also that of future collective rights.

A more refined version of the collective-rights argument than Weiss’s is offered by Edward Page (2006: 150-158). He distinguishes the reductionist view of collective rights from the holistic view. The reductionist view takes the rights of a group to be simply the aggregation of the rights of its members. In contrast, the holistic view, which he prefers, maintains that a group’s rights are not reducible to the individual rights held by its members. The assumption here is that a group is an entity bearing ethical status independently of its individual members. It seems that the holistic view, unlike the reductionist one, can respond to the non-identity criticism. Page concedes, however, that when a current course of action harmful to a certain future group is a necessary condition of that group’s coming into existence, the non-identity problem of the group arises. He therefore concludes that the holistic view is a partial solution, which resolves the problem only in some cases.

True, there are cases in which the argument of future collective rights holistically understood seems to have solidity, while that of future individual rights fails to do so. Suppose that 10 percent of the whole population of a

¹There are several attempts to justify reparations for historical atrocities without invoking the notion of intergenerational rights. One of these attempts is the argument from inheritance. This argument notes that the past Africans were morally entitled to compensations made by their contemporary European wrongdoers. It is then contended that the Africans of today inherit, from their ancestors, the entitlements to compensations. However, the inheritance argument encounters the problem of the temporal boundaries of historical liability (Usami, 2007: 165). George Sher (2005: 191-200) suggests a more sophisticated response to the non-identity problem; however, his response also seems susceptible to the temporal-boundary problem.
country will vary 20 years later, depending on which environmental policy the government chooses at present. In this case, the individual-rights claim collapses as a result of the contingent identity of these varying individuals, whereas the collective-rights claim may survive because the identity of the whole population will be roughly maintained. However, the holistic view can have some force insofar as it concerns immediately succeeding generations. In cases of distant non-overlapping future generations—say, those living in the country 200 years later—the identity of the entire population will radically depend on the current policy choice and events thereafter. Page is right when he admits that the holistic view is merely a partial solution, but the range of the solution might be more limited than he expects.

When it is put into the context of proposed international reparation, the idea of collective rights faces the perplexing question of who counts as a member of the rights-holder group. Let us take Mathieu’s case again. One might say that all citizens of the Republic of Benin share the collective right to require the Portuguese and the French to compensate them for the historical slave trade. If this is correct, Mathieu appears to have this right. The question arises, however, of who really holds the collective reparatory right. Thanks to the enormous profit from slave trading, the Kingdom of Dahomey prospered in that land and its ruling ethnic group, the Aja, accumulated great wealth. If Mathieu is a current member of the Aja, does he count as a legitimate rights-holder? Does he rather fall within a group of wrongdoers? If some of his ancestors were among the Aja and others belonged to an oppressed group, the Fon, is he a victim or an offender? The notion of collective reparatory rights does not work in solving the real-world issues of compensations for historical injustices, until current members of the victimized group are identified.

Threshold Conception of Harm

Another form of attempts to defend the idea of intergenerational rights against the non-identity criticism bases this idea on the threshold conception of harm.\(^1\) Meyer (2003: 143-144, 147-159) addresses this conception in discussing both posterity’s rights and reparatory justice. He begins by noting that according to liberal cosmopolitanism, future people have general rights vis-à-vis us. However, cosmopolitans take the view that currently living people cannot have just claims to compensations for historical injustices committed against their predecessors. He rejects the latter view and purposes to provide the common grounds for posterity’s rights and compensatory claims. For this purpose, he advances what he calls the subjunctive-threshold interpretation of harm, the formula of which reads: ‘Having acted in a certain way (or having refrained from acting in that way) at a time \(t_1\), we thereby harm someone only if we cause this person’s life to fall below some specified threshold’ (Meyer, 2003: 147). Meyer concedes that the scope of the subjunctive-threshold

\(^1\)Jeff McMahan (1998: 223-229) discusses several issues related to the threshold understanding of harm.
interpretation of harm is limited, and thus he finally suggests the combination of this interpretation and the more traditional one.

Meyer and Dominic Roser (2009) develop the sufficientarian theory of currently living people’s duties and future people’s correlative rights, which is based on the threshold conception of harm. The position they particularly advocate is strong sufficientarianism—as opposed to weak sufficientarianism—which gives absolute or lexical priority to the improvement in the well-being of those who live below the threshold, and which holds that the worse their level of well-being, the more it matters that they receive benefits. Meyer and Roser maintain that one of justifications for intergenerational sufficientarianism is that this view successfully responds to the non-identity problem.

Given the threshold reading of harm, a future person can be said to be harmed by us when our choice of actions and policies causes her life to fall below some reasonable threshold of harm, even if this choice is the necessary condition of her coming into existence. Similarly, some currently living people can be described as harmed by the ancestors of others when the ancestors’ wrongdoings made their present welfare fall below the threshold. Putting aside the complex question of what a reasonable threshold means, these implications are readily understandable.

The most serious defect in Meyer’s argument is that the threshold understanding of harm does not succeed in defending the notion of intergenerational rights against the non-identity objection. Suppose that our choice of the policy of environmental destruction will cause $P_D$ to live below the specified threshold. Should $P_D$ have the right to an above-threshold life, this right paradoxically would demand the nonexistence of $P_D$. The same point can be made in relation to the alleged reparatory rights. If John holds the right to compensations for historical slavery because he lives below the threshold, then this right presupposes the entitlement to live beyond the threshold. In a counterfactual world where the slavery did not occur, however, John cannot enjoy an above-threshold life because he does not come into existence in the first place. The threshold conception of harm fails to protect intergenerational rights from the non-identity problem in the contexts of concern for posterity and compensatory justice.

**Perfect and Imperfect Duties**

My discussion in the previous sections was intended to establish that several notable versions of the rights theory ultimately provide no foundation for intergenerational concern or reparatory justice. After examining some objections to the rights of future people and rejoinders to them, I explicated how the non-identity problem threatens both the notion of posterity’s rights and that of the right to compensations for historical wrongs in domestic and international settings. Next, it was shown that the collective-rights argument

---

1Meyer and Roser (2009: 233-236) try to define what the threshold of harm means.
has no force in cases of distant successors and that, as to compensatory justice, this argument faces the question of who counts as a current member of the victimized group. Then, I argued that the threshold conception of harm fails to defend the idea of intergenerational rights against the non-identity criticism.

As I have tried to show, the non-identity problem is too robust for upholders of intergenerational rights to dispose of. If I am right about this, a puzzle arises: Why do so many scholars of ability struggle for such a hopeless project? What lies behind their persistent attempts to defend the idea of rights in intergenerational relationships against the non-identity objection?

At least a part of the answer is, I suspect, the ambiguity of a traditional dichotomy between perfect and imperfect duties (Usami, 2006: 271). The perfect/imperfect distinction about duty or obligation, which dates back to the Stoics, notably Cicero, is notoriously ambiguous. From Ambrose to Aquinas to Kant, classical authors have interpreted this distinction in a great variety of ways. Among the items on a long list of different meanings of the dichotomy are culpable/supererogatory, important/trivial, specific/vague, legal/nonlegal, and correlative/non-correlative.1 It has long been recognized that the duty not to kill falls into the category of perfect duty, while the duty of charity is a paradigm case of imperfect duty. Yet a few typical examples do not help to clarify meanings of that distinction because the examples are susceptible of various depictions. The obligation of charity is often taken as the supererogatory nonlegal recommendation that correlates with no right of the receivers of charity.

Given the ambivalence of the perfect/imperfect dichotomy, one may assume that, when a duty of one party does not correlate with any right of the other party, the duty must be supererogatory, trivial, vague, and nonlegal. Under this assumption, if one believes in the culpability or importance of some duty, one may feel that one needs to demonstrate its correlation to some right. Theorists of intergenerational rights seem to be trapped in such an obsession. Their conviction that our obligation to respect the interests of future people is significant drives them to contend that these people have rights vis-à-vis us. Because they are convinced that affluent citizens in Europe bear the moral duty to contribute to improving the well-being of those living in sub-Saharan Africa, they feel compelled to assert that the latter group hold compensatory rights vis-à-vis the former.

The non-correlativity of a duty, however, does not logically imply its supererogatory character, trivialness, vagueness, or nonlegality. This point is well illustrated by a familiar example. My duty to pay taxes does not correlate with any specific right of other citizens. Despite its non-correlativity, this duty is neither supererogatory nor trivial, and it is specifically provided by relevant tax laws. Once one distinguishes among different meanings of the perfect/imperfect dichotomy, one can conceive the idea of a non-correlative duty without worrying about the unintended degradation of the duty. Obligation to take account of the interests of future generations falls within the

---

1 The ambiguity of the perfect/imperfect dichotomy is concisely described and decomposed by Millard Schumaker (1992).
category of non-correlative duty, and the obligation is arguably justified with no reference to posterity’s rights.¹

With regard to reparative justice, it stems from the non-identity problem that only the survivors and members of the victim’s family who were born before the misdeeds occurred have a compensatory right, which correlates to a duty of the perpetrators or their government. Injustice must be rectified promptly. The limited temporal scope of compensatory rights does not indicate that the descendants of those victimized by historical injustices do not matter in a society or the international community. They do matter. In a domestic setting, a variety of policy options should be considered and pursued for the sake of minorities and indigenous peoples, which might include anti-discrimination laws, affirmative action programs, multicultural policies of language and education, and political arrangements for self-government. Severe poverty in sub-Saharan Africa and other parts of the world deserves high priority on the agenda of the international community. Governments of affluent societies, transnational institutions, and nongovernmental organizations have many things to do in order to improve the well-being of the global needy.²

Advocates of intergenerational rights have struggled against the non-identity problem. However, their attempts so far have not been successful, as I argued. This problem rather seems to suggest that we should seek some approach other than the rights theory.

References


¹I discuss elsewhere our intertemporal obligation from the perspective of intergenerational fairness (e.g., Usami, 2006: 272-279).

²My own view of global justice is based on the human right to subsistence, and it focus on the current realities of international political economy instead of the distant past (Usami, 2007: 165-169; Usami, 2008a).


