

AN ANTHOLOGY OF PHILOSOPHICAL STUDIES VOLUME 4

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

Chrysoula Gitsoulis
CUNY
USA

Jonas S. Green
Birkbeck College
UK

Keith Green
East Tennessee State University
USA

Jan-Christoph Heilinger
Berlin-Brandenburg Academy of Sciences and Humanities
Germany

Mark McEvoy
Hofstra University
USA

Chris Onof
Birkbeck College
UK

Elizabeth Schiltz
College of Wooster
USA

Andrew Ward
University of York
UK

**Athens Institute for Education and Research
2010**

**AN ANTHOLOGY OF
PHILOSOPHICAL STUDIES
VOLUME 4**

**Edited by
Patricia Hanna**

First Published in Athens, Greece by the Athens Institute for Education and Research.

ISBN: 978-960-6672-65-1

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 ATINER SA.

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

©Copyright 2009 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 <i>Patricia Hanna</i>	1
Part 1: History of Philosophy		
<i>Ancient and Medieval</i>		
2.	Aristotle versus Plato on Tragedy with Respect to Civic Education <i>Jaroslav Danes</i>	7
3.	The Practice of Dialogue: Socrates in the Meno <i>J. Gregory Keller</i>	19
4.	Lugones, Agency under Oppression, and the Aristotelian Practical Syllogism <i>Crista Lebens</i>	27
5.	The Examined Life: Outline of a Neo-Socratic Argument <i>Joe Mintoff</i>	37
6.	The Milindapanha ('Questions of King Milinda'): An Early Encounter between Greek and Buddhist Philosophy <i>Jere O'Neil Surber</i>	47
7.	Plato on Truth and Falsehood in the Polis <i>Anne M. Wiles</i>	59
<i>Modern</i>		
8.	Locke on Personal Identity: The Form of the Self <i>Siyaves Azeri</i>	69
9.	How Cartesian is Cartesian Linguistics? <i>Christina Behme</i>	79
10.	The Conceptualist Analysis of Descartes's Creation Doctrine <i>Daniel Considine</i>	89
11.	Hegel and Foucault Re-visited <i>Evangelia Sembou</i>	99
Part 2: Metaphysics, Logic and Philosophy of Language		
12.	The Individual Variability Problem <i>Dimitra Electra Gatzia</i>	113
13.	How to Avoid Relativism: What Forms the "Bedrock" that turns our Spade? <i>Patricia Hanna</i>	129
14.	Two Dimensions of Logical Theory: Commitment and Truth <i>John Kearns</i>	141

15.	In what Sense is Tarski's Semantic Conception of Truth Semantic?	153
	<i>Ladislav Koren and Jana Pavlikova</i>	
16.	Being as the Set of All Things	167
	<i>Donald V. Poochigian</i>	
17.	The Self's Survival and its Freedom	177
	<i>Andrew Ward</i>	
18.	Ramsey and Laconicism	187
	<i>Kiriake Xerohemona</i>	
Part 3: Ethics and Political Philosophy		
19.	Sartre and Judith Butler: A Problem of Agency That Existentialism Can't Solve	201
	<i>Katherine Cooklin</i>	
20.	The Social Foundations of the Ethical in Kierkegaard's Idea of Existence	211
	<i>Michael J. Matthis</i>	
21.	Responsibility in an Age of Global Climate Change: Lessons from the Phenomenology of Thomas Langan	221
	<i>Ingrid Leman Stefanovic</i>	
22.	From Rationality to Morality: The Challenge of Egoism	229
	<i>James P. Sterba</i>	

List of Contributors

Siyaves Azeri is a Lecturer at Department of Philosophy, University of Ottawa, ON, Canada. His areas of interest and research include Modern Philosophy, philosophy of science, consciousness, subjectivity, and philosophical psychology. He has written articles in English, Turkish, and Persian; his papers have appeared in book volumes and several journals including *Filozofia* and *Science and Society*.

Christina Behme is PhD Candidate at the Philosophy Department at Dalhousie University in Halifax, Nova Scotia, Canada. She also has a degree in marine biology (Diplombiologe) from the University of Rostock, Germany. Her main research interests are in Philosophy of Language, Philosophy of Science and Evolutionary Theory. She coordinates the multidisciplinary Evolution Study Group at Dalhousie and has recently organized a module on language evolution. She has published on language acquisition and language evolution and is currently completing her PhD thesis on Cartesian linguistics and language acquisition.

Daniel Considine is Visiting Assistant Professor of Philosophy at Metropolitan State College in Denver, Colorado, USA. He teaches history of modern philosophy, philosophy of science, and logic. His main research interests lie in early modern European philosophy, especially Descartes, Hume, and Kant, epistemology, and philosophy of science.

Katherine Cooklin is Assistant Professor of Philosophy at Slippery Rock University of Pennsylvania, USA; she teaches Social and Political Philosophy, Philosophy of Law and Justice, and Phenomenology and Existentialism. She received her Ph.D. in philosophy from the University of Texas at Austin, and she specializes in Social and Political Philosophy from a Continental perspective. She has published papers on Julia Kristeva and Michel Foucault, and her current research interests include Continental and feminist models of subjectivity within the field of Philosophy of Law.

Jaroslav Danes is Associate Professor at the Faculty of Philosophy of the University of Hradec Kralove, Czech Republic. He teaches the History of Ancient and Modern Political Thought. He works on the history of political thought and science (especially medicine). He has published several articles in Czech and English: his latest authored work is *Is the conception of the origins of medicine in De prisca medicina original?* (In: *Listy filologické* 1-2/2008), and his most recent published book is *John Philoponus versus Aristotle on the Eternity of the World* (Prague, Czech Academy of Science Publishing 2006). He is currently working in the field of Political Thought in Ancient Greek Tragedy.

Dimitria Electra Gatzia is Assistant Professor of Philosophy at the University of Akron, Wayne College, USA. Her research interests include the metaphysics of color, color perception, and how cognitive psychology and neuroscience can inform our philosophical theories of the relation between perception and

cognition. She has published on color fictionalism, the individuation of the senses, and the role of the ethics of care and economic theory.

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.

John Kearns is Professor and Chair of the Philosophy Department of the University at Buffalo, the State University of New York, USA. He has published a number of books and articles on logic and philosophy of language. These include *Using Language: The Structures of Speech Acts* and *Reconceiving Experience: A Solution to a Problem Inherited from Descartes*, both published by SUNY Press. Some of his articles have appeared in the *Journal of Symbolic Logic*, the *Notre Dame Journal of Formal Logic*, *History and Philosophy of Logic*, *Logique et Analyse*, and *Linguistics and Philosophy*. His current research is focused on speech act theory and the logic of speech acts.

J. Gregory Keller is Senior Lecturer in the Department of Philosophy at Indiana University-Purdue University, Indianapolis, Indiana, USA, and Research Fellow, Center for the Study of Religion and American Culture. He teaches Twentieth-Century Philosophy, Ethics, and Philosophy of Culture. His research interests include dialogue as life practice, philosophy of literature, cosmopolitanism, philosophy of religion, the relationship between morality and thought, and concepts and practices of utopia. His papers have appeared in such journals as *Philosophy and Literature*, *ProtoSociology: An International Journal of Interdisciplinary Research*, and *Sophia: International Journal for Philosophy of Religion, Metaphysical Theology and Ethics*.

Ladislav Koren is Assistant Professor at the Philosophy Faculty of the University Hradec Králové, Czech Republic, where he teaches philosophy of language, logic, epistemology and political philosophy. He has written a number of articles and reviews in Slovak and English. He is a member of LMS centre (an interdisciplinary research centre for Language, Mind and Society). His research interests lie mainly in philosophy of language, logic and epistemology.

Crista Lebens is Associate Professor of Philosophy in the Philosophy & Religious Studies Department at the University of Wisconsin-Whitewater, USA. Feminist Philosophy, Philosophy of Race and Gender, and Classical Philosophy number among the classes she teaches. Her work centers primarily on the social ontology of gender and race and has appeared in *International Studies in Philosophy* and in the anthology, *The Center Must Not Hold: White Women Philosophers on the Whiteness of Philosophy*.

Michael J. Matthis is an Associate Professor of Philosophy at Lamar University in Beaumont, Texas, USA. He teaches courses in existentialism, philosophy of art, and philosophy of religion. He has written papers on

Kierkegaard, Nietzsche, and aesthetics, has read papers at conferences both nationally and internationally, and has been published in *Philosophy Today*, *New Scholasticism*, and *Man and World*. He has also edited a collection of essays in a book, titled *The Beautiful, the Sublime, and the Grotesque: The Subjective Turn in Aesthetics from the Enlightenment to the Present*, to be published this year by Cambridge Scholars Publishing.

Joe Mintoff is a Lecturer in Philosophy at the University of Newcastle, Australia. His primary research interests are in moral philosophy, specifically the theory of rational choice, the philosophy of Socrates, and ancient approaches to the question of how to live. His articles have appeared in *Australasian Journal of Philosophy*, *American Philosophical Quarterly*, *Ethics*, and *Ratio*.

Jana Pavlikova is Assistant Professor at the Philosophy Faculty of the University Hradec Králové, Czech Republic, where she teaches didactics of philosophy. She is the author (or co-author) of several articles on didactics of philosophy published mainly in Czech journals. She is a member of LMS centre (an interdisciplinary research centre for Language, Mind and Society).

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, forthcoming publications include articles entitled “Human Nature and Human Rights” and “An Epistemological Critique of Computer Learning.” Forthcoming presentations include articles entitled “Mathematical Identity: The Algebraic Unknown Number and Casuistry” and “An Economic Paradox: Entropy and Growth.”

Evangelia Sembou holds a D.Phil. from the University of Oxford. Her thesis is on G. W. F. Hegel’s *Phenomenology of Spirit*. She has published articles on Hegel in *History of Political Thought*, the *Jahrbuch für Hegelforschung* and the *Bulletin of the Hegel Society of Great Britain*. She has taught political theory and philosophy at the University of Oxford and philosophy at Birkbeck College, University of London, UK.

Ingrid Leman Stefanovic is a Professor of Philosophy at the University of Toronto, Canada. She has served as founding Director of the University’s interdisciplinary Centre for Environment. She has held posts such as Executive Co-Director of the International Association for Environmental Philosophy, and Chair of the Steering Committee of the United Nations University Toronto Regional Centre of Expertise on Education for Sustainable Development. Her research interests and publications are in the area of ethics, as well as environmental and architectural phenomenology. A recent book is entitled *Safeguarding Our Common Future: Rethinking Sustainable Development* (SUNY 2000) and an edited volume on *The Natural City: Re-Envisioning the Built Environment* is forthcoming.

James P. Sterba is Professor of Philosophy at the University of Notre Dame, USA. His previous publications include *Affirmative Action for the Future* (Oxford, 2009), *Does Feminism Discriminate Against Men? – A Debate*

(Oxford, 2008 – co-authored with Warren Farrell), *The Triumph of Practice over Theory in Ethics* (Oxford, 2005) and *Terrorism and International Justice* (Oxford, 2002) and *Justice for Here and Now* (Cambridge, 1998). He is past president of the American Philosophical Association (Central Division) and several other organizations.

Jere O’Neill Surber is Professor of Philosophy and Cultural Theory at the University of Denver, USA. He has been a visiting professor at such institutions as the Johannes Gutenberg Universität/Mainz, the Katholieke Universiteit/Leuven, the University of Tromsø, Norway, and Oxford University, UK. He has published several books and numerous articles on 19th and 20th century Continental Philosophy and on contemporary cultural theory and critique. He also teaches Asian and comparative philosophy and has several recent articles forthcoming in these areas. He is currently completing a major work entitled *What is Philosophy?: Embodiment, Signification, and Ideality*.

Andrew Ward teaches Philosophy at the University of York, UK. He has also taught at the Open University, UK and at the University of Florida at Gainesville, FL, USA. He writes mainly in the areas of Personal Identity, Aesthetics, and the Philosophy of Immanuel Kant. Publications include: *Kant: The Three Critiques* (Polity Press: Cambridge, 2006) and articles in philosophical journals, including *Mind*, *American Philosophical Quarterly*, *Philosophy and Phenomenological Research*, and *History of Philosophy Quarterly*.

Anne M. Wiles is Professor of Philosophy at James Madison University in Harrisonburg, Virginia, USA. She teaches ancient Greek philosophy from the pre-Socratics to Plato and Aristotle. She has published articles on the philosophy of Plato, Aristotle, Rabindranath Tagore, Dante, St. Thomas Aquinas, and Jacques Maritain, and a chapter in *Plato and Platonism*, “Predication in the Later Dialogues of Plato” (CUA Press, 2006). Her current research centers on the application of Gadamer’s hermeneutical method to the discussion of unity and plurality in Parmenides, Plato, Aristotle, and Thomas Aquinas.

Kiriake Xerohehmona is a Lecturer at Florida International University, USA. She teaches Logic, Symbolic Logic, Philosophy of Logic, and Language and Paradox. She works in the field of Analytic Philosophy. Her research interests include philosophy of logic, metaphysics, and ethics. Other research interests include Philosophy of Deception, and Pragmatism. She has received an award for Excellence in Teaching and since 2006 she is the Director of the Master of Liberal Arts Program.

CHAPTER ONE

Introduction

Patricia Hanna

This volume is a collection of papers selected from those presented at the 4th International Conference on Philosophy sponsored by the Athens Institute for Research and Education (ATINER), held in Athens, Greece at the St. George Lycabettus Hotel, 1-4 June 2009.

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. In addition to the presentations, in 2009 we were able to provide opportunities for special interest discussions; these included a discussion of new directions in philosophical research, and the use of philosophy to teach writing in English to non-native speakers (ESL writing) across a variety of disciplines in the sciences. We plan to include more of these in the future, and are looking into ways of providing progress reports in future publications as appropriate.

Over the course of the conference, sixty papers were presented. The twenty-one papers 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 approaches 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.

Patricia Hanna
Editor
USA

Part 1

History of Philosophy

Ancient and Medieval

CHAPTER TWO

Aristotle versus Plato on Tragedy with Respect to Civic Education¹

Jaroslav Danes

The institution of dramatic performances had an immense political importance in classical Athens, which is plain from both the references of contemporaries and the research of modern scholars.² The organisation and operation of theatrical performances were a public activity connected to the state and the apparatus of democratic Athens. Luciano Canfora³ says that dramatic performances were one of the three pillars holding up the Athenian democracy. The playwrights, as well as politicians (it is difficult to separate the two categories), saw in drama an instrument of civic education and probably also a form of political discussion.⁴ Therefore, it is not surprising that Plato and Aristotle, as two of the most original and brightest figures of Athenian intellectual dissent⁵ and as clever critics of the Athenian democracy, paid particular attention to tragic poetry as one of the most important democratic institutions. In this article we would like to answer the question of whether, according to Plato and Aristotle, it is possible and useful to educate citizens in the theatre and by tragic pieces.

Plato and Tragedy

Tragic poetry and its function in society occupy Plato's mind most in three discernible political dialogues, *Gorgias*, *the Republic*, and *the Laws*. In all three works his attitude to tragic poetry is damnatory and critical. Plato's objections to tragedy form a part of his comprehensive criticism of democracy, covering metaphysical, socio-psychological, and ethical considerations. In order to grasp the gist of his rejectionist position we, then, have to consider briefly Plato's objections to democracy, which focus on the democratic view of education and the democratic set of values.

¹This article was supported by the Czech Science Foundation Research Programme No. 401/09/P032, Greek Tragedy and Political Thought. All quotations are borrowed from the Loeb classical library (see bibliography); abbreviations follow the Liddell & Scott & Jones (1996)

²I mean especially The Great Dionysia. See Isocrates, *De pace* 82; Demosthenes, *De corona* 120; Goldhill (1990: 97-129; 2000: 60-65); Meier (1993: 44-51).

³Canfora (1994: 124)

⁴Aristophanes, *Frogs* 1054-1055; *Acharnians* 496-509.

⁵We have borrowed this category from Ober (1998).

Plato's analysis was based on an assumption, or rather an analogy, between the individual and society. In his opinion "the soul" of society is a projection of individual souls (*Rep.* 435c)¹. Each and every soul consists of three capacities: reason, spirit, and appetite. The chief and the strongest element is the appetitive capacity, 'because of the intensity of its appetites concerned with food and drink and love and their accompaniments'² The appetitive part in each soul naturally predominates. Its hypertrophy threatens both the individual soul and society. The other two parts are appointed to watch over the appetitive part and the rule is entrusted to reason (*Rep.* 442a). For the appetite is manifold. In the course of exposition Socrates draws a distinction between necessary appetites and unnecessary ones (558d-559d). The necessary appetites are, in modern parlance, subsistence needs. We cannot get rid of them. Moreover, we profit by satisfying them. Let us take food as an example. The necessary appetite is to eat something; on the contrary, the unnecessary appetite is to overeat. Describing the change from the oligarchic character to the democratic character, Socrates says

*'And in the end, I suppose, they (unnecessary appetites) seize the citadel of the young man's soul, finding it empty and unoccupied by studies and honourable pursuits and true discourses, which are the best watchmen and guardians in the minds of men who are dear to the gods....And they themselves prevail in the conflict, and naming reverence and awe thrust it forth, a dishonoured fugitive. And temperance they call want of manhood and banish it with contumely, and they teach that moderation and orderly expenditure are rusticity and illiberality, and they combine with a gang of unprofitable and harmful appetites to drive them over the border.'*³

This process is supported by the drunkenness of liberty. Plato characterises democracy as a regime totally dedicated to liberty and equality. Socrates' definition of liberty in the *Republic* is the same as Pericles' definition in the *Funeral Oration*⁴: to live as you like. But whereas Pericles circumscribes the limits of liberty (law, unwritten laws etc.), Socrates caricatures this view of liberty, identifying liberty with arbitrariness and anarchy. With respect to the concept of the soul and the thesis that politics and society are a projection of individual characteristics, Plato denounces the democratic concept of liberty as dangerous because of its failure to support the cultivation of the appetitive part of the soul and its not promoting the supremacy of reason. According to Plato, the democratic concept of liberty contains a self-destructive dynamics which implies anarchy or tyranny⁵. If man lives as he likes and does what he wants,

¹Schofield (2006: 268-269) demonstrates the incorrectness of the analogies drawn by Plato between the individual and society.

²*Rep.* 580d-e

³*Rep.* 560b-e; Cf. *Lg.* 855c.

⁴Thucydides 2.37

⁵Cf. Schofield (2006:107-112; 117-121)

he is, in Plato's opinion, identical to a tyrant since a tyrant is distinctive about his arbitrariness. At the same time such an arbitrary soul is claimed to be slavish, because nothing prevents it from blind obedience to appetites and desires.¹ This irrational inclination is, according to Plato, a universal human dream (*Leg.* 687c). We can only oppose this inborn inclination if we pursue a specific education. The elements of this so-called *cathartic* education are most properly accounted for in the *Sophist* (227d-229c). The Eleatic Stranger minutely elucidates what the soul has to be purified of, what it means to purify one's soul, and how a soul should be purified. A soul should be purified of everything which differs from excellence (*ἀρετή*) and of what is evil. There are two kinds of evil described in the dialogue. The first kind of evil is defined as a disorder of the soul and a lack of temperance. The second one is ignorance. Whereas the first kind lies within the competence of the law, the second one should be removed by *cathartic* education based on *elenchos* (analytic thought and demonstration). *Elenchos* helps us to discard false beliefs and to promote shame, originating in the reflection of one's ignorance. It is, then, philosophy which lies at the heart of *cathartic* education or the noblest kind of rhetoric, as Plato probably might have said.

Plato defines tragic and comic poetry as an imitation (*μίμησις*) of characters pursuing various practices². Plato's evaluation of imitation is quite complicated. Generally speaking, mimesis is rejected from the epistemological point of view. At the same time he makes a distinction between good and bad imitation in the field of education. Why was imitation rejected? Imitation is not, by definition, cognitive activity. In the *Republic* (601c-602b) he exemplifies the difference in the case of the flute. Socrates evaluates and speaks about the knowledge or art of the maker, user, and imitator. The flute-player knows perfectly which flute serves rightly for flute-playing, because generally the user of anything is the one who knows the most about the things which are used. The maker will have to believe the user. But the imitator does not know if a thing is rightly or badly made and will imitate a thing that appears beautiful to the ignorant mass. The imitator does not have either knowledge or the right opinion on the things that are imitated. Yet if his imitation supplies pleasure to the mass audience he will go on imitating the flute without knowledge of the flute. Imitation is thus a kind of play without any claim to serious activity aiming at knowledge and truth. Imitation and its forms (painting, poetry) are the art of illusion. They occupy a third place, remote from reality and truth (*Rep.* 602c).³ The truth is available only to dialectics, which is an instrument of the rational part of the soul (*Rep.* 532a-534e). On the other hand, in the discussion about the foundation of the state governed by philosophers Socrates describes this state as an imitation of the divine pattern and compares philosophers to painters (*Rep.* 501b). In the *Laws* the Athenian Stranger says in reply to the question of what would happen if

¹Cf. The final passage of *Alcibiadēs* (135a-e). Though the authenticity of this dialogue has been disputed for hundreds of years, the opinions contained in the dialogue are truly Platonic.

²*Rep.* 394b-c ; *Lg.* 655e, 817b

³See 597b-e.

strangers were to put on tragic performances in the perfect state, that the lawgivers of the perfect state would be the authors of the best and the most beautiful tragedy because their constitution is an imitation of the best and the most beautiful life – the life of the gods (817b).¹

Plato was not consistent in his rejection of imitation and drew a distinction between good and bad imitation.² He introduced and employed this difference both in the *Republic* and in the *Laws*. But what is the definition of good or bad with respect to imitation? In the *Republic* Socrates gives an account of good and bad imitation in connection with the musical education of the guards. By finishing his scrutiny of what and how to say about the gods, heroes, and *daimones*³ Socrates moves on to an inquiry into the human area and the plausibility of literal styles. The guards should concentrate on the perfection of their souls and personality, and therefore they should not tackle the common arts and imitation. If they are to imitate something, they must imitate only what pertains to their character and function in society: freedom, rationality, moderation, temperance, piety etc. (395c). On the contrary, they should throw off the imitation of heart-breaking and emotionally exalted behaviour, no matter if we mean joy or sorrow. In other words, the imitation of behaviour which can be classified as good is that of a disciplined, moderate, brave and prudent man.

In this point Plato's social psychology, pedagogy and political thought coincide in the form of a criticism of democracy, including the institution of dramatic performances. The tragic poet does not focus on the rational part of citizens' souls and even cannot influence it, because if he wishes to be successful, he must supply pleasure to the mass audience (*Rep.* 605-607). Being under the influence of the audience and being obliged to arouse pleasure, the tragic poet is a slave to the audience and the audience plays the role of a tyrant (*Grg.* 502b). This role appears to be the same as the role of an orator in the assembly or the defence lawyer in the court. They all, as Socrates puts it in *Gorgias* (502c), speak to a mass audience, trying to please it. The only difference lies in the fact that the tragic playwright addresses the audience in verses or meter and his performance is coloured by music. But what would remain if we ignore the metrical form of the verses? Only sober speech would remain. This is the reason why Socrates considers tragedy to be a branch of assembly speech and his criticism of tragic poetry transforms itself into a criticism of political oratory. So, the masses govern not only the state (the assembly and courts), they even dominate in the theatre. In *Laws* (701a) this situation is described as theocracy. Theocracy appears to be a reason why

¹On the basis of these considerations we might argue that Plato's dialogues are a good imitation.

²In the *Republic* Socrates (392d-394c) scrutinises three kinds of styles, aiming to complete by this scrutiny the account of that what one should speak about and how one should speak: pure, imitative, and combined. He gives examples of these styles. Dithyrambic poetry represents the pure style, the tragedy and the comedy are described as imitative, and finally, epic poetry is an example of the combined style. When Socrates makes an evaluation of these styles he absolutely banishes both the tragedy and the comedy.

³*Rep.* 379a-392c.

all attempts to educate citizens in the theater are futile. Whenever a playwright addresses the audience critically, he arouses strong opposition, disappointment, disapproval, and even whistling from the audience. Therefore it seems unreasonable both to expect and to promote shame, respect for wiser opinions and wiser people, and fear in the audience. On the contrary, we are justified in expecting wicked insolence (ἡ πονηρὰ ἀνασχυντία), which is caused by exceedingly shameless freedom (ἐλευθερίας λίαν ἀποτετολμημένης)¹. Thus, tragedy is doomed, like other democratic mass institutions, to be a slave to the delight and mind of the masses. Tragedy is not a means of education; rather it appears to be a generally accepted source of amusement and flattering rhetoric.² With respect to our account of cathartic education, we are justified in saying that tragic poetry is an enemy of philosophy because it promotes disorder and ignorance in the souls of citizens and makes itself servile to the appetitive part of the soul. However, it would be more correct to argue that the problem does not lie in the tragedy itself but in the democratic audience, which it addresses. The opposition of tragic poetry to philosophy is the opposition of the mass democratic institution to elite activity, rhetoric to *cathartic* education, and rationality (*elenchos*) to irrationality (persuasiveness). With respect to Plato's opinion on the role of philosophy in society, we also see the opposition of the private to the public. Cathartic education, being individual and not widely available, lies in the sphere of the private.³ To pursue philosophy in a democratic society is both impossible and dangerous. As Schofield persuasively demonstrates, in *Gorgias* and the *Republic* Plato reformulated the Socratic attitude to the role of philosophy in society with respect to democratic Athens.⁴ Whereas in the *Apology* Socrates sees himself as someone whose philosophical activity means a kind of public service and the good of the populace of Athens (36c-d) and his death proves that philosophy is unimaginable outside democratic Athens, in the *Republic* Socrates appears to be someone who is not engaged, who prefers a private life to a public one, and who scorns the unlawful deeds of other citizens whose behaviour can be compared to that of wild animals, which makes each and every form of philosophy impossible (496a-e).

¹*Lg. 701b*. To be correct we have to mention that the passage of *Laws* does not pertain specifically to tragic poetry, but generally to the rules of musical art. However, these rules also cover tragic poetry.

²In *Minos* (321a) Plato uses the following characterisation for tragedy: the most persuasive and delightful to the mass (δημοτερέστατον τε καὶ ψυχαγωγικώτατον).

³Salkever (1990:284) talks about Plato's and Aristotle's preference for private over public education. I am not convinced of this preference in the case of Aristotle with respect to the eighth book of *Politics*. However, in Plato's dialogues we find the same means that are also common in tragedy: metaphors, imitations, myth etc. But the opinion on the essential impossibility of educating the masses underlies Plato's analyses.

⁴Schofield (2006:19-30).

Aristotle on Tragedy

Aristotle's work can be perceived in many ways as opposing Plato's. The same might be said in the case of the educative function of tragedy. However, Aristotle shares many theoretical presuppositions with Plato. In accord with Plato, he presupposes that the majority of people prefer a disorderly life to a life of excellence or virtue (*Pol.*1319b31-34, *EN* 1168b15-20). In consonance with Plato he assumes this inclination to be something which generally pertains to human biological nature – all beings naturally seek pleasure, which appears to be the satisfaction of their needs and natural appetites. On the contrary, Aristotle does not share Plato's prejudices and bias against the middle and lower classes. Aristotle draws a distinction between those who are serious about living virtuously (*σπουδαῖοι*) and those who aim at no definite end, but at *pleonexia* (*φᾶῦλοι*) in a form of disorderly desire for money, honours, or sex (*EN* 1168b15-20).¹ We can try to resist *pleonexia* in three ways: laws, education, and philosophy. (*Pol.*1267b) In Plato's opinion the mass is immune to all three means. Aristotle objects to his view in the fourth book of *Politics* (1297a7-13), where defending the mixed constitution, he argues that the *pleonexia* of the masses subverts the regime and constitution less than the *pleonexia* of wealthy citizens. Therefore it is not wise to create a purely aristocratic constitution. We can see an essential disagreement with Plato in the question of the masses or populace as the collective body making political decisions. Discussing the legitimacy of power and questions of sovereignty, Aristotle tries to give an answer to the question of whether the leading position in the state belongs to the masses (*τὸ πλῆθος*) or a few high-born and excellent men (*τοὺς ἀρίστους μὲν ὀλίγους δέ*). Aristotle argues that though the individuals of the populace need not be people who are serious about living virtuously (*σπουδαῖος ἀνὴρ*), as a collective body they are excellent and in possession of extraordinary moral and intellectual faculties which even surpass the faculties of a virtuous man. For this reason the collective civic body is expected to be the better judge not only of politics but also of music and poetry.² On the other hand, Aristotle is cautious and relativises his arguments with the statement that it is not clear if his thesis of the wisdom of the collective body is true for every populace. Let us, then, see, how he evaluates populaces in different types of democracies. Though partly in consonance with Plato's low opinion of democracy and favouring *politeia* (*Pl.* 1295a25-1296a22) under the present circumstances, nevertheless Aristotle considers democracy to be the best of the bad regimes (*Pl.* 1289a26-30). The best form of a bad democratic regime appears to be that in which peasants and herdsmen dominate, since they have average property and must work intensively, in consequence of which they restrain themselves from the inclination towards power and greed which characterises the majority of people (*Pl.*1318b7-18).

¹*Pleonexia* literally means having more. This category occupies a central place in the moral and political thought of antiquity. See Balot (2001).

²*Pol.* 1281b-1282b

The worst form of democracy seems to be that in which the masses govern and where citizenship is allowed even to the poor and to persons without an orderly civic origin in order to make the masses prevail in numbers over the middle class and aristocrats (*Pl.*1319b). In this type of regime everybody is allowed to live as he likes and to do as he wants. In effect, there will be licence among citizens and even among children, women, and slaves. This type, then, is similar to tyranny, because both a tyrant and a citizen are characterised by sheer arbitrariness. In this point Aristotle's opinion on the collective body of the bad form of democracy converges with Plato's opinion on democracy and the character of a democratic citizen in the eighth book of the *Republic*. Is it futile or possible to temper the tyrannical dream of the masses, who are not open to philosophy and rational argumentation? Is not the populace open to any other form of persuasion or amelioration? I think that Aristotle's ideas on tragedy go in this direction, which is the opposite of Plato's.

In *Poetics* Aristotle gives the following definition of tragedy:

*'Tragedy, then, is the mimesis (μίμησις) of an action which is elevated, complete, and of magnitude; in language embellished by distinct forms in its sections; employing the mode of enactment, not narrative; and through pity (ἐλέου) and fear (φόβου) accomplishing the catharsis (κάθαρσιν) of such emotions.'*¹

Aristotle's definition of and the following treatise on tragedy have always appeared puzzling to scholars and readers. For in *Poetics* we are left neither an account of the importance of pity and fear nor an explanation of the concept of *catharsis*. Besides, Aristotle says that each art is aimed at something. In *Poetics*, however, the aim of tragedy is very briefly sketched without any specification. Aristotle says that tragedy aims at arousing pleasure which comes out of pity and fear (1453b10-13). In the passage where Aristotle makes preparations for the account of the plot as the soul of tragedy, we are told that the plot is what is the most persuasive and most influences the soul of the spectator (τὰ μέγιστα ψυχαγωγεῖ).² Then the aim of tragedy is to persuade and influence the souls of the spectators. In this point tragedy coincides with rhetoric. The opinions on tragedy expressed by Aristotle are similar to those of Plato, who characterised tragedy as a branch of rhetoric. For the poet has to use the same means as an orator has to. Therefore, in *Poetics*, while explaining the problem of diction and thought, Aristotle refers to his discourses in *Rhetorics* (1403a36). But the poet and the orator differ in one substantial detail. The handling of events in tragedy must have an effect even without a direct statement, while the orator produces an effect by delivering a speech. Aristotle mentions the coincidence of tragedy with rhetoric one more in *Poetics*, while analysing the importance of thought or the 'capacity to say what is pertinent and apt, which in formal speeches is the task of politics and rhetoric' (1450b).

¹*Po.* 1449b23-28. Halliwell (2005:49)

²1450a33

According to Aristotle, the older poets made characters speak like politicians, while contemporary poets make them speak like orators (1450b5). The persuasive faculty, which is the aim of the tragedy, is also common to rhetoric and political science, for they all consider the question of a good and bad enactment. Political science, in contrast to tragedy, is a discursive discipline, while tragedy, similarly to a special kind of rhetoric, uses examples and metaphors.

Who or which audience might be moved by tragic plays? Who and how should the tragedy persuade? In opposition to Plato, Aristotle does not think that only one absolute method of explanation and persuasion exists. In *Metaphysics* (995a1-15) Aristotle speaks about the relationship between the kind of audience and the character of the explanation. Some people accept only mathematical explanations, while others do not have the faculties to follow a minute and consequent line of arguments, which makes them prefer explanations using examples or they are used to listening to myths. In *Rhetorics* Aristotle argues that the listeners to rhetoric are those who are not able to *'follow a lengthy chain of argument'* (1357a). According to Aristotle, the majority of people lack such a faculty. However, it does not mean that something good and true cannot be revealed to them.

From this we move on to Aristotle's treatise on pity and fear. Aristotle scrutinises these emotions in *Rhetorics*, while exposing the influence of the emotions on judgement.¹ There he puts the definition of fear as a *'painful or troubled feeling caused by the impression of an imminent evil that causes destruction or pain'* (1382a). The things that are to be feared are those which arouse pity. What is to be feared most is the errors which *'once [they] have been committed it is impossible to repair'* (1382b). However, not all people are liable to fear.

*'Those who either are, or seem to be, highly prosperous do not think they are likely to suffer anything; wherefore they are insolent, contemptuous, and rash, and what makes them such is wealth, strength, their number of friends, and power. It is the same with those who think that they have already suffered all possible ills and are coldly indifferent to the future...for it is a necessary incentive to fear that there should remain some hope of being saved from the cause of their distress. A sign of this is that fear makes men deliberate...'*²

So, fear is an emotion which rationalises decision-making and action.

Pity is also *'a kind of pain excited by the sight of evil, deadly or painful, which befalls one who does not deserve it; an evil which one might expect to come upon himself or one of his friends, and when it seems near'* (1385b). As in the case of fear, not all people are liable to pity. The group of such people is almost of the same kind as in the case of fear. On the contrary, women,

¹*Rh.* 2.5, 2.8.

²*Rh.* 1383a.

children, and old and educated people are an ideal audience for arousing pity. Thus, pity and fear are interconnected and they cannot both be expected from the two poles of the populace.

The notion of *catharsis* in connection with art and education can be found, apart from *Poetics*, only once in Aristotle. We mean a passage in *Politics* (1341b39), where Aristotle thinks about the importance of musical education, which, apart from amusement and education, can also lead to *catharsis*. Aristotle promises that he will return to the exposition of this notion in his treatise on poetry. However, this return is not extant.¹ Hundreds of years of research have brought us three different interpretations of *catharsis*, which are summarised by Salkever (1990: 292)²:

1. religious expiation (*Po.* 1455b);
2. homeopathic treatment of emotional disorder in the manner of Hippocratic medicine (*Pl.* 1341a-1342a);
3. clarification or enlightenment.

None of the three interpretations appears to be apt for Aristotle's definition of tragedy. For neither of these interpretations poses the question of who is the object of catharsis with respect to fear and pity. Paradoxically, Aristotle himself neglected this question in *Poetics*. We are in accord with Salkever, that the interpretation of catharsis depends on '*the best characterisation of the nature of the audience tragedy specifically addresses*' (1990:292). Surely tragedy does not address an audience which is not sensible to pity and fear (see above). The ideal audience for tragedy is the majority of the populace (οἱ πολλοί), those who stand between the extremes mentioned by Aristotle in his treatise on fear and pity. What means does tragedy use to purify fear and pity? Pity and fear '*arise above all when events occur contrary to expectations yet on account of one another. The awesome will be maintained in this way more than through a show of chance and fortune*' (1452a). Fear and pity are not aroused in the case of an arbitrary character. Aristotle says that the character must be of an average position, '*not preeminent in virtue and justice, and one who falls into adversity not through evil and depravity, but through some kind of error (ἀμαρτία)*' (1453). What does Aristotle mean by error (ἀμαρτία)? It appears to be human fallibility based on ignorance. However, Halliwell warns us that we should not exclude even deeds of passion and limited culpability on the character's part.³ According to Nussbaum, the notion of error (ἀμαρτία) covers various forms of faults, which, however, are not implied by the agent's

¹As Stephen Halliwell says, this reference to poetry might pertain to the lost book(s) of *Poetics* or to the lost Aristotle dialogue *On Poets*. In all these books Aristotle may have explicitly introduced *catharsis* to block the Platonic charge that the arousal of emotions by tragedy tended dangerously to increase susceptibility to the same emotions in life. (2005: 18)

²Salkever refers to Lucas (1968: 273-290). Cf. Halliwell (1986: Appendix 5).

³Halliwell (2005:17)

(corrupted) character.¹ Lucas (1968: 302) considers error to be ignorance in connection with a lack of bad intention, which Aristotle himself explains in the *Nicomachean Ethics* (1110b)². What area does the error affect and what are the consequences which arouse pity and fear?

‘What tragedy must seek are cases where sufferings occur within a relationship (ἐν ταῖς φιλίαις), such as that of brother and brother, son and father, mother and son, son and mother – when the one kills (or is about to kill) the other, or commits some other such deed.’³

The notion of relationship (φιλία) is one of Aristotle’s ethical and political keys (*EN* 8). Aristotle works with φιλία in two meanings: family relationships and civic cohesion (*EN* 1161b). Friendship and love are, on the one hand, an element of the family, on the other an element of civil society (*EN* 1155a; *Pl.* 1262b). Friendship (and love) seems to be the necessary condition for a happy and meaningful life (*EN* 1151a5). A plot which arouses pity and fear by destroying the familial φιλία hits the most sensitive point of the citizens’ souls and it is the gist of the tragic art. The plot of the tragedy most probably afflicts the mass audience and common citizens who have families and children.

In this point we can see the educative function of the tragedy. Tragic *catharsis* makes citizens more rational and sensitive to good and bad behaviour. The means of the *catharsis* are apt with respect to the nature of the audience because understanding deeds which damage family ties is, in contrast to philosophy, available to almost everybody. From this point of view the tragedy has shown itself to be the ideal means of educating a democratic citizenry. If this interpretation of Aristotle’s works is correct, then the tragic *catharsis* is a functional analogue of the Platonic *cathartic paideia*. The theory of *catharsis* may thus be understood as a response to Plato’s criticism, which is also indicated by the cross-references to Plato’s *Republic* in the key passages of *Poetics*, *Rhetorics*, and *Politics*.⁴

¹Nussbaum (2001: 378-395). Nussbaum refers to detailed analyses of the notion of error (ἀμαρτία). See especially Stinton (1975), Bremer (1969), and Dave (1967).

²Srv. Else (1957: 376-80) and Salkever (1990: 297).

³*Po.* 1453b18-21.

⁴Aristotle’s critical attitude to Plato is highlighted at the end of the second book of *Poetics*, where Aristotle makes a comparison of tragic with epic poetry. We saw Socrates’ preference for epic poetry to tragic in the *Republic*. On the contrary, at the end of *Poetics* (1462b) Aristotle supplies arguments which favour tragic drama over epic poetry. ‘Tragedy achieves its effect even without actors’ movement, just like an epic... Add the fact that tragedy possesses all an epic’s resources as well as having a substantial role for music and spectacle, which engender the most vivid pleasures. Again tragedy has vividness in both reading and performance. Also, tragedy excels by achieving the goal of its mimesis in a shorter scope... Also, the imitation of epic poets is less unified.’ (1462b)

Conclusion

Plato denounced tragedy as a mass democratic institution corrupted by the tyrannical inclination of the audience. Tragedy appeared to him to stand in opposition to philosophy, reason, and the cultivation of the irrational parts of the soul. Plato was convinced of the impossibility of public education. He perceived tragic poetry as an anti-educative means, which strengthened the democratic drunkenness of freedom and nurtured tyrannical inclinations in citizens' souls. Generally speaking, a democratic mass institution promotes democratic thinking, which is the reason why no amelioration and education can be reached. Indeed, we also find passages in Aristotle's works concerning the corrupting influence of a bad audience on the tragic poets¹. However, Aristotle rejected Plato's bias against the audience. First, in Aristotle's opinion the audience as a civic collective body has a collective intelligence, and second, the audience is open to persuasion and cultivation in the form of emotional *catharsis*. By arousing pity and fear tragedy corrects the misperception of freedom and happiness as a tyrannical dream. Thus tragic poetry is a proper form of democratic civic education, which cultivates the emotions contributing to rational decision-making.

Bibliography

- Aristophanes (2002). *Frogs*. Cambridge: Harvard University Press.
Aristophanes (1998). *Acharnians*. Cambridge: Harvard University Press.
Aristotle (2005). *Poetics*. Cambridge: Harvard University Press.
Aristotle (2006). *The Art of Rhetorics*. Cambridge: Harvard University Press.
Aristotle (1933). *Metaphysics*. Cambridge: Harvard University Press.
Aristotle (1934). *Nicomachean Ethics*. Cambridge: Harvard University Press.
Aristotle (1932). *Politics*. Cambridge: Harvard University Press.
Balot, R. (2001). *Greed and injustice in classical Athens*. Princeton: Princeton University Press.
Bremer, J. M. (1969). *Hamartia: tragic error in the Poetics of Aristotle and in Greek tragedy*. Amsterdam.
Canfora, L. (1994). *Storia della letteratura Graeca*. Roma: Laterza.
Dawe, R. D. (1968). *Some Reflection on at and hamartia*. Harvard Studies in classical Philology 72: 89-123.
Demosthenes (1926). *De corona*. Cambridge: Harvard University Press.
Else, G. F. (1957). *Aristotle's Poetics*. Cambridge: Harvard University Press.
Goldhill, S. (1990). *The Great Dionysia and Civic Ideology*. In: Winkler, J. J. & Zeitlin, F. I., *Nothing to do with Dionysos?: Athenian drama in its social context*.

¹*Poetics 1453a*. This fragmentary criticism in *Poetics* finds support in the passage of *Rhetorics* (1403b) where Aristotle scrutinised the persuasiveness of orators in connection with their way of presentation. In this passage Aristotle makes complaints about the baseness of both contemporary tragic performances and politics. However, Aristotles' criticism is not meant as criticism of the educational function of tragic poetry itself.

An Anthology of Philosophical Studies

- Goldhill, S. (2000). *Athenian Drama and Political Thought*. In: Rowe, Ch. & Schofield, M., *Cambridge History of Greek and Roman Political Thought*. Cambridge: Cambridge University Press.
- Halliwell, S. (1986). *Aristotle's Poetics*. London: Duckworth.
- Halliwell, S. (2005). Introduction. In: *Aristotle, Poetics*. Cambridge: Harvard University Press.
- Isocrates (1929). *On the peace*. Cambridge: Harvard University Press.
- Liddell, H.G. & Scott, R. & Jones, S. (1996). *Greek-English Lexicon*. Oxford: Clarendon Press.
- Lucas, D. W. (1968). *Aristotle Poetics*. Oxford: Clarendon Press.
- Meier, Ch. (1993). *The political art of Greek tragedy*. Oxford: Polity Press.
- Nussbaum, M.C. (2001). *The Fragility of Goodness: Luck and Ethics in Greek Tragedy and Philosophy*. Cambridge: Cambridge University Press.
- Ober, J. (1998). *Political dissent in democratic Athens: intellectual critics of popular rule*. Princeton: Princeton University Press 1998.
- Plato (2006). *The Republic*. Cambridge: Harvard University Press.
- Plato (1926). *Laws*. Cambridge: Harvard University Press.
- Plato (1925). *Gorgias*. Cambridge: Harvard University Press.
- Plato (1914). *Apology*. Cambridge: Harvard University Press.
- Plato (1921). *Sophist*. Cambridge: Harvard University Press.
- Salkever, S. (1990). *Tragedy and the Education of the Dēmos*. In: Euben, P., *Greek Tragedy and Political Theory*, 274-304. Berkeley & London: California University Press.
- Schofield, M. (2006). *Plato: The Political Philosophy*. Oxford University Press.
- Stinton, T.C. (1975). *Hamartia in Aristotle and Greek Tragedy*. *Classical Quarterly* 25: 221-254.
- Thucydides (1919). *History of the Peloponnesian War I*. Cambridge: Harvard University Press.

CHAPTER THREE

The Practice of Dialogue: Socrates in the Meno

J. Gregory Keller

As I attempt to address this issue I find myself already caught in a web of multiplicity so vast and dense that I despair of finding any path through it. Let me then begin with a general claim and two sets of points by which we might orient ourselves. The claim: dialogue might be pursued either under the aspect of sameness and self or under the aspect of difference and other. My interest in the topic of dialogue covers both of these approaches, and each is, I believe, worthy of pursuit. They are even, one might say, hard to distinguish at their further points, like straight lines that continue parallel to infinity but that appear to converge at the horizon of our view. I offer two categories of points of reference: (1) specific pursuits of dialogue as life practice: (a) seeking truth and (b) moral endeavor, and (2) two modes of pursuit: (a) listening and speaking and (b) acting without self-defense.

Let us listen first to words from a well-known twentieth-century thinker, Robert Frost, who tells us: ‘Something there is that doesn't love a wall’ (1915, line 1). The practice of dialogue begins with this fact; it always calls us to note that no building of walls, whether epistemological or social or of mind or heart, remains the final word in our engagement with life. Later in *Mending Wall* Frost's narrator says:

*‘Before I built a wall I'd ask to know
What I was walling in or walling out,
And to whom I was like to give offence.
Something there is that doesn't love a wall,
That wants it down.’* (1915, lines 32-36)

I think that Socrates would approve, with the added proviso that that which one risks offending would be truth.

Let us turn now to Socrates, and to a particular text. I do not claim to be either revealing or discussing the historical Socrates, a task too arduous and perhaps simply impossible to accomplish in these pages. I address instead the Socrates of one particular Platonic dialogue, the *Meno* (1924), recognizing that current scholarship generally places the *Meno* at a turning point in the Platonic corpus,¹ a beginning of the move away from the more Socratic early dialogues

¹See e.g., Ionescu (2007) pp. xi-xii and Scott (2006) pp. 6-7.

to the middle period in which it is believed that Plato describes his own mature views.

Much of interest occurs in the *Meno*; we will take up only two strands of that multiply-threaded tapestry: the discussion concerning whether virtue is desire of the ‘fine’ or beautiful and the brief suggestion that progress is being made when one discovers that one is perplexed. I want to address these pieces of the *Meno* merely schematically, with the hope that further dialogue might flesh out the discussion from the perspective of other interlocutors.¹

Desire for the good

In the dialogue bearing his name, Meno has had considerable trouble answering the seemingly simple question of Socrates about the nature of virtue. Socrates insists first of all that neither he nor anyone he has heard knows the true definition of virtue, to which Meno expresses amazement. When, however, Meno attempts to give an answer to the question as to the nature of virtue, he finds it perplexingly difficult. We join the dialogue at the point at which Meno tries again, basing his new claim on a bit of literature. It seems that a poet has, conveniently, said that virtue is a desire for and the power to attain the beautiful (*kalon*). Meno quite approves of this definition and draws from it his current definition of virtue: virtue is a matter of the right desire and the right power. Socrates immediately takes up this claim, dividing it at its mid-point, so as to address for now only its first conjunct: virtue is desire for the beautiful.

We might briefly summarize Socrates’ argument as follows: To desire the beautiful is to desire the good. Some then, Meno insists, desire the good and some the bad (or ugly!). Some desire the bad thinking it is good; others desire it recognizing that it is bad. Of those who desire the bad knowing it to be such, some think it will bring them benefit; those of course, Socrates points out, desire the bad as though it were good, at which point he briefly offers a definition of the good as that which benefits and of the bad as that which harms. If there remain some who desire the bad knowing it is bad and recognizing that the bad harms, they must recognize also that to desire something is to possess it. These people then desire to possess that which harms (them); in other words, they desire that which will make them miserable and unhappy. But this latter group seems to have no members, for, Socrates says and Meno agrees, no one wants to be miserable and unhappy. This line of argument then brings the discussion to a conclusion – the conclusion that no one desires the bad knowing what they are doing. The interlocutors offer at this point a major amendment of the original claim: it cannot be that what distinguishes the virtuous from others is desire for the good; everyone desires the good.

We might note in passing that the text makes a subtle but possibly important move from speaking of ‘desire’ (*epithumein*) to speaking of ‘want’

¹For helpful recent discussions about the *Meno*, see, e.g., Ionescu, 2007; Scott, 2006; and Weiss, 2001.

(*boulesthai*); in the end we are told that no one ‘wants’ evil, not that no one desires it.¹ Whether this distinction makes a difference to the argument is difficult to discern. Perhaps the difference is this: People do in fact *desire* things that are evil, even believing that these things bring harm, misery, and unhappiness; but no one *wants* evil in the sense of wanting the harm, misery, and unhappiness itself. This claim could be restated as: even though *S* does not want *y* (misery, let’s say), *S* can still desire *x* (the thing that produces misery); this seeming paradox relies on the quite obvious difference between implication in fact (*x implies y*) and recognition of that implication in one’s desires. I can, in other words, desire something without at the same time desiring its obvious consequences. The status of ‘wants’ might be, as Aristotle suggests, such that ‘without qualification and in reality, what is wished [wanted, *boulesthai*] is the good, but for each person what is wished is the apparent good’ (III, 4, 1113a, 24, 2000, p. 37). Want/wish then is a term for the aim at the good that orients all one’s deliberations and choices (Aristotle again); desire is a term for pure appetite. Whatever we wish to make of these distinctions, Socrates has clearly steered Meno away from his initial apparently elitist belief that he (Meno) and his friends are the only ones who are virtuous due to their desire for fine or beautiful things to a more egalitarian recognition that everyone desires or at least wants the good (and therefore, of course, the beautiful).

The Value of Perplexity

In this brief bit of dialogue, Socrates takes a conclusion Meno proposes (Meno having failed in two previous attempts to define virtue) and shows that on its own terms it is unsuccessful. Having dismissed the one conjunct, Socrates then takes on the other conjunct concerning the power to acquire good things. This discussion shows, primarily, once again the shallowness of Meno’s understanding. He deems the good to be wealth, political power, and political honors. Socrates leads him to see the faultiness of his aims. We will not follow that argument further here but will instead move past this discussion to the attack Meno levels at Socrates. In its simplest terms it reveals the frustration with philosophy as a project, a frustration that most of us experience at times and that students, especially beginning students, appear to feel regularly. If questions are the essence of philosophy (as many of us would maintain²), then what about answers? Must we dismiss answers entirely as irrelevant? Meno’s approach is more specific of course; he complains that he has moved from a comfortable state of being able to give speeches on virtue to a distinctly uncomfortable state of being speechless about the very same topic. The perplexity to which he has been reduced, he suspects, arises from tricks by which Socrates has enchanted him. The complaint then is not only that Meno

¹See e.g. Ionescu, 2007, pp. 25-27, for further discussion of this distinction.

²In this regard one can hardly do better than look at the delightful last chapter of Russell (1997) in which he describes questioning as the essence of philosophy’s contribution to human life.

dislikes the feelings he now has; it is that these feelings results from some sort of illegitimate trickery on the part of the philosopher.

The simplest answer to Meno's complaint is to say that he is an arrogant, and not very smart, apprentice railing against the wise master of the trade (of philosophy) who has merely shown him up for the fool he is. We might say that his perplexity heralds the possibility of wisdom. 'It is,' we might tell Meno, 'darkest before the dawn.' Just stick around, my friend, and you will learn; your perplexity will vanish in the light of truth; and so on. Unfortunately, although Socrates does argue that perplexity proves to be a step in the right direction, we do not in the end see questions answered and the longing for truth and wisdom fulfilled¹ (as if these are the things for which Meno longs). At the very end of the dialogue Socrates mentions a possible though still uncertain conclusion on the basis of their reasoning so far then says but for now we need to stop our inquiry only to return to these issues another day.

This means that there is here both some good news and some bad. The good news is that those of us who consider ourselves philosophers, even in the most limited ways, will not soon be out of a job – philosophy continues because questions continue, apparently without end. The not-so-good news is that the answers we may have wished for in engaging in this form of inquiry just keep receding further and further away. One might begin the study of philosophy (as I did) with the aim of eventually getting it right (and I, personally, didn't think it would take that long). Experience in the field, however, leads one to conclude that a lifetime might, under the best of circumstances, only get one started properly understanding the questions.

Turning back to the *Meno*, the analogy of the slave boy learning something new in geometry offers an important point – when he is perplexed he is better off than when he falsely thought he knew. At the same time it can offer false hope about the kind of issue Meno and Socrates are discussing. Although mathematics is by no means as simple and straightforward as it might have seemed to the Greeks and to those of us with only an elementary understanding of it, it does offer, if one keeps to certain principles and draws very clear boundaries, some interesting and obvious answers to questions. Philosophy, on the other hand, suffers from the great advantage of being a 'community of the question' (Derrida, 1978) for whom every answer only tells us what question to ask next. This is not to say that we cannot, by keeping to certain principles and drawing very clear boundaries, proclaim ourselves to possess some answers. It is only to say that once out of our philosophical infancy we recognize that the principles and boundaries are not a solution but another part of the problem; they are themselves always up for question.

¹Yet is not wisdom the result of beginning to know what one does not know?

Teaching Virtue

Let us turn then from Meno's perspective, that of the confused and angry recipient of too many questions, to that of Socrates. What is he doing? The initial question of the *Meno* is 'Can virtue be taught ...?' to which Socrates answers with more than a touch of irony, what an amazing question – here I am not even knowing what virtue is and you have leaped beyond this level of ignorance to another arena of thought entirely by asking whether it can be taught. Socrates does not allow the question to stand of course but moves the discussion immediately into an attempt to address the definition of virtue. It turns out then that Meno does not know what he claims to know; he does not know what virtue is. After the interlude during which Socrates questions the slave boy about geometry, they return to the question of whether virtue can be taught and Socrates reluctantly agrees to face this question directly. In the end, however, the question receives only a provisional answer, for how can it be answered until one knows what virtue is?

Some would claim,¹ and I am among them, that Socrates is answering the question all along the way. Meno begins with a high level of un-virtuous, perhaps even vicious, arrogance. Socrates, in clearing the ground of apparent knowledge plants seeds of virtue in Meno's soul. Even if wisdom is not the heart of virtue (a topic addressed with uncertain results in the *Meno*), one can hardly be virtuous without it. And the heart of wisdom, it would seem even to those of us without much claim to wisdom itself, is a recognition that we know far less than we believe we do. The logic of question and answer² is also a morality. One cannot truly ask without humility. One cannot stand face to face with a genuine interlocutor without allowing for at least reciprocity between oneself and the other and perhaps allowing for something more than reciprocity.³ One cannot face the Other – whether in the form of an idea, a question, or a person – without opening oneself to the profound realization of one's own lack and the world's mystery and, in some form or other, the possibility of transcendence.

So Socrates here questions and in questioning answers, though not in the only way, the question he faces. Can virtue be taught? First, we do not know what virtue is and therefore must remain open in the face of the question and all its progeny. Second, we can only approach such a question not with a set of absolutes but with an earnest and continuing inquiry. Third, as with life in general, we can bring to our aid a set of curious possibilities and odd metaphors, following the twisting and obscure course of discourse itself, tracing the river of thought to its source, which is never exactly a source, since

¹See e.g., Scott (2006), esp. chapter 16, 'Meno's progress.'

²Gadamer (1991) pp. 370-372 aligns his discussion of the logic of question and answer with R. G. Collingwood (1978) and (2002).

³Levinas (1998) follows Buber (1996) in pushing the importance of dialogue past 'experience' to 'encounter' and then on to show that genuine dialogue takes us past any sense of control over the other or even reciprocity with the other to a sense of debt to the other. Levinas marks out the ethics of our encounter with the other as 'first philosophy' in other texts, e.g., 1969.

the river is not only impossible to enter once (or twice¹ – even if we translate ‘river’ into the metaphor of river stages following Quine, 1961), it has a source which could be called both itself and not itself in the endless, unrepeatable cycle of water flowing to the seas and the seas returning to the sky and the sky returning to the rivers. Fourth, and not finally for in this endeavor there is no end, we must not only speak our answers and listen to the questions they contain; we must note the way life answers and questions, speaks and listens, and tune our tin ears to the music emanating from the process itself in all its harmonics and varieties and discordant melodies.

Dialogue as Moral Practice

I return now to say that the web of the multiplicity of questions and answers, and of answers and questions, lies just beneath the surface of our everyday dialogues as well as of our hearty philosophic ones. Socrates stands at our side whispering words of encouragement when we falter and of discouragement when we think we have succeeded. The thing one must love about philosophy (whether one will or not) is that it offers wonder in place of certainty, questions to every answer, and a long endeavor that shows us exactly why in thought as well as fact:

*Something there is that doesn't love a wall,
That sends the frozen-ground-swell under it,
And spills the upper boulders in the sun,
And makes gaps even two can pass abreast.* (Frost, 1915, lines 1-4)

Let us return for a moment to issues with which we began: the aspects of sameness/self and difference/other and the undertaking of dialogue as seeking truth and as moral practice through modes of listening-speaking and of countering self-defense. Dialogue as practiced by Socrates might seem to remain within the aspect of sameness/self in that Meno has no voice once he has provided his series of faulty definitions or his attack on Socratic perplexity. The alleged dialogue of reason has been accused of being merely a hegemonic monologue, that is, a single voice listening only to itself and beating all foreignness into submission.² Yet if Socrates is, as he claims, perplexed himself and if he pursues the question of the nature of virtue in earnest, the process of question and answer becomes a movement guided not by self and not monotonically, but listening always to the other of ‘reason’. This other might not be a simple sameness but might itself involve an encounter with a

¹There’s some confusion as to what Heraclitus says about stepping into rivers, but one traditional claim is that he says that no one can step into the same river twice (as Plato has it in *Cratylus* 402a, but compare Heraclitus fragment DK22B12). Cratylus, then, is supposed to have said in response that one cannot even do so once (Aristotle *Metaphysics* 1010a13, where the Platonic claim about Heraclitus is repeated).

² See e.g., Levinas (1969) and (1998).

beyond that cannot be contained within any single, simple perspective. Reason itself might be less a choir without discord than a symphony full of contrasts, disharmonies, multiplicities, and difference (yet with guiding melodies of a rich complexity not ever to be wholly contained by human thought). In that case, listening to reason and attempting to speak with or for it becomes an encounter rather than a monologue.¹

We see, then, in the Socratic practices described above a truth-seeking dialogue as moral practice as well as modes of listening and speaking. The place, finally, of countering self-defense must be addressed here briefly. Meno finds himself in this dialogue constantly defending himself, but against what we might ask. It seems that he defends himself against the truth, but how might such a defense benefit him?

Are we, one might ask, benefited by maintaining our version of the truth against all questions? Are we made better, our lives made richer, by turning aside from that which questions, from that which would say nay to our certainties? Or are we instead benefited by refusing, finally, to defend ourselves against truth and questions, benefited by allowing walls to tumble to the ground instead of maintaining them at all costs? Socrates, and here I hope we would follow him explicitly, says we are only benefited by examining our lives and thoughts, our ideas and ideals, until we welcome genuine perplexity and shun miserly certainty, making the logic of question and answer a moral stance that leads us to see that virtue can at least be taught to the extent of our recognizing learning and questioning themselves as virtues and the refusal to learn or question as a vice.

The problem of self-defense lies in identifying ourselves with our opinions. If I believe that my beliefs of the moment stand not merely as one set of ideas I hold at one time to be replaced in the course of things by others, I might so identify with them that I take a given set of beliefs to be me. When that is the case I may defend them with the same level of violence, at least on the verbal level, with which I would defend my life. Socrates calls his interlocutors to recognize their interests as lying on the side of truth rather than on the side of the maintenance of current beliefs. Thus the self-defense one might engage in to uphold or rebuild the walls of belief can turn instead to open acceptance of the potential to learn. Self-defense, in terms of unquestioning defense of one's beliefs, becomes the enemy of truth and Socratic dialogue the enemy of self-defense.

So in our philosophical wandering and wondering, we see that walls stand, sometimes, for the simple purpose of reminding us to ask and wonder and question and encounter that which we think we know but do not. And when we meet a wall that does not want to be met in return, we must join the earth in the disassembly of that wall or at least we must join one another in raising the question by which it is, perhaps, truly reinforced, until, that is, Socrates comes along and asks the question we have missed, producing a gateway through the

¹Gadamer (1991) addresses such a possibility in his discussion of the I-Thou relation, pp. 358-362.

crumbling barricade and an open passageway, where at least two may walk and talk, and where no path existed before.

Bibliography

- Aristotle. (2000). *Nicomachean ethics*. (T. Irwin, Trans.) Indianapolis: Hackett Publishing.
- Buber, M. (1996). *I and thou*. (W. Kaufmann, Trans.) New York: Touchstone Books.
- Collingwood, R. G. (2002). *An essay on metaphysics*. New York: Oxford University Press, USA.
- Collingwood, R. G. (1978). *An autobiography*. Oxford, England: Oxford University Press.
- Derrida, J. (1978). *Writing and difference*. (A. Bass, Trans.) Chicago: University of Chicago Press.
- Frost, R. (1915). 'Mending Wall.' In: *North of Boston*. New York: Henry Holt and Company.
- Gadamer, H. G. (1991). *Truth and method*. 2nd revised ed. (J. Weinsheimer & D. G. Marshall, Trans.) New York: Crossroad.
- Ionescu, C. (2007). *Plato's Meno: An interpretation*. Lanham, MD: Rowman & Littlefield Publishers.
- Levinas, E. (1998). 'Dialogue: Self-consciousness and the proximity of the neighbor.' In: *Of God who comes to mind*. (B. Bergo, Trans.) Stanford, CA: Stanford University Press.
- Levinas, E. (1969). *Totality and infinity*. (A. Lingis, Trans.) Pittsburgh, PA: Duquesne Press.
- Plato. (1924). *Meno*. In *Plato with an English translation: Laches, Protagoras, Meno, Euthydemus*. (W. R. M. Lamb, Trans.) Loeb Classical Library. London: Harvard University Press.
- Quine, W.V.O. (1961). Identity, ostension, and hypostasis. In: *From a logical point of view*. 2nd ed. Revised. Cambridge, MA: Harvard University Press.
- Russell, B. (1997). *The problems of philosophy*. New York: Oxford University Press.
- Scott, D. (2006). *Plato's Meno*. Cambridge: Cambridge University Press.
- Weiss, R. (2001). *Virtue in the cave: Moral inquiry in Plato's Meno*. Oxford: Oxford University Press.

CHAPTER FOUR

Lugones, Agency under Oppression, and the Aristotelian Practical Syllogism

Crista Lebens

If, while reading a philosopher whose work one knows well, one comes across an unusual reference to another philosopher's work, there can be three explanations for the seeming anomaly: first, it was a tangential reference that does not add much of significance to the analysis or argument, second, it is a problematic reference that overall does not support the argument, and third, it is a useful reference that adds a great deal to the argument. In her articulation of Marilyn Frye's concept of arrogant perception, María Lugones makes reference to Aristotle's concept of the practical syllogism. Given that the usual body of work that a woman of color feminist such as Lugones draws on typically does not include Aristotle, I found this reference intriguing.

Since philosophical work is done 'in conversation' with other thinkers, one can locate a general body of work that a philosopher is likely to engage in developing her theories. Typically Lugones draws on radical lesbian women of color feminists such as Gloria Anzaldúa and Audre Lorde, white radical lesbian feminist philosophers such as Marilyn Frye and Sarah Hoagland, and Latin American thinkers such as Walter Mignolo. To find a reference to Aristotle, then, is a bit puzzling, perhaps anomalous. I found it intriguing enough to pursue the use Lugones would make of Aristotle. What I found was not just a passing reference. Nor was it a use of Aristotle that, given his views on women and slaves, ultimately undermined the strength of the analysis. In fact, Lugones makes a quite useful connection to Aristotle, though she does not make as much of it as could be done.

This appropriation of Aristotle is useful not simply because he offers a concept of practical reason that supports Lugones' overall project, but because his problematic statements about women and slaves serve as an example of arrogant perception and can lead to a deeper understanding of gender, race, and class relations. Lugones works out the former application; in this paper, I will articulate the latter connection.

Frye on Arrogant Perception

Marilyn Frye defines oppression as a network of social forces and barriers that simultaneously press and restrict the oppressed person's choices. The lives of oppressed agents become structured around suiting the interests of the

oppressor(s). (Frye 1983a 2-4). Frye's account explains how, for example, a woman is indoctrinated in feminine values from birth onward, and then "chooses" options that serve the interests of men. Using the term 'arrogant eye' or 'arrogant perceiver' to denote social expectations of women, especially in relation to men, Frye shows how arrogant perception appropriates the woman's energy and interests. In turn, the woman monitors herself according to these norms.¹ The way out of this structure according to Frye is through the gaze of another, the 'loving eye,' who recognizes the independence of the woman, the beloved, and authorizes the beloved to pursue choices on her own. (Frye 1983b 72-74) Lugones reads this analysis of oppression and subordination through an Aristotelian lens. Her use of the practical syllogism further articulates the dynamic between the woman and the arrogant perceiver.

The Practical Syllogism

In Book VI of the *Nicomachean Ethics*, Aristotle defines a prudent (or practically wise) man as one who 'deliberates well concerning what is good and expedient for himself, for living well' (*NE*, 1140a 26-31). John Cooper argues that the practical syllogism is not a stage in deliberation, (33) but is the link by which decision leads to action (46). Thus for Aristotle, practical reason, which leads to action, is distinguished from theoretical reason, which leads to propositions or beliefs. Deborah Modrak describes the practically wise person as one who is 'good at a particular kind of means-end reasoning, deliberation, where the starting point (the end) is a moral principle and the conclusion is a decision about what should be done here and now' (Modrak 1994, 211).

In contrast, Cooper describes the weak-willed or incontinent person (*akratēs*) as one who:

Does not stand by what he has deliberated, but abandons it, whereas the strong man holds to what he has deliberated. . . . The weak and the strong do not differ as to proairesis [a commitment to act based on desire²], the choice of action that they have arrived at by deliberation, but the one acts against this proairesis whereas the other manages to act as he has decided he should (Cooper, 47).

¹"The arrogant perceiver . . . coerces the objects of his perception into satisfying the conditions his perception imposes. He tries to accomplish in a glance what the slave masters and batterers accomplish by extended use of physical force, and to a great extent he succeeds. He manipulates the environment, perception, and judgment of her whom he perceives so that her recognized options are limited, and the course she chooses will be such as coheres with his purposes" (Frye 1983b, 67).

²See Cooper, 47 n. 59. Aristotle's discussion of weak and strong willed persons can be found in book VII 1145b10-12, 1150b19-21, 1152a18-19, 27-29; see also I 13 1102b14-18, IX 4 1166b8-10, 8 1168b34-35, [*Eudemian Ethics*] II 7 1223a37-38 (Cooper 47 note 58).

Both the strong- and the weak-willed person know the right thing to do, but the weak-willed person, overcome by desire or appetite, does not act on that knowledge.

Regarding Aristotle's account of the reasoning of slaves and of free or citizen women and men, he claims that one who is a slave 'by nature' has no ability to deliberate, (*Politics*, 1260a 13-14) yet can recognize the reasoning of the master, and so can understand and follow orders (*Politics*, 20-23). Citizen women, according to Aristotle, have the ability to reason, but this ability 'does not have authority' (*Politics*, 1260a 8-14). According to Modrak, Aristotle thinks women can reason about right action, i.e., can know what to do, but do not put that knowledge into practice, therefore, women on Aristotle's model are weak-willed (Modrak 1994, 213). Furthermore, Modrak argues that Aristotle, contrary to Plato, believes this is not a culturally-induced deficiency in reason, but one that exists by nature (Modrak 214-215). Citizen men's ability to reason is by nature unqualified, according to Aristotle; therefore these men are suited to rule women and male and female slaves.

The Practical Syllogism and Arrogant Perception

Aristotle recognizes that the woman has the power to reason, specifically the power to deliberate about action. In Lugones' interpretation, there are two practical syllogisms at work when a woman under arrogant perception deliberates: the arrogant perceiver's and the woman's (Lugones 56). Uncoerced, she might act on her own reasoning. However, under arrogant perception, the choices she makes are directed to suit the interests of the arrogant perceiver. This analysis fits with Modrak's argument that, according to Aristotle, women's weakness of will must be bolstered by the guidance of men. Though Modrak does not use the same language, she could be talking about the power of arrogant perception which constructs the *akratic* person to be one who conforms readily to the expectations of the arrogant perceiver.¹ The way out of this construction, according to Lugones, is to recognize that one exists in 'worlds' of meaning where her values are not constructed by arrogant perception.

Lugones uses the practical syllogism to show that the woman can authorize her own actions without relying on a beloved other, as in Frye's model. She argues that one is subordinated to the arrogant eye in some worlds of meaning, i.e., some socially constructed realities, but not in others:

¹“What a morally weak person needs to enable him or her to be virtuous is some way to make the morally correct, general principle that *ex hypothesi* is accepted by the agent sufficiently vivid in the context of action so that he or she will recognize its implications for his or her behavior. This would be achieved were a practically wise individual present to remind the akratic person of the principle and its implications. Understanding marriage as a relation between a person who is capable of deliberation and one who is much less adept at it, Aristotle might be led to conclude that the way to avoid failures of moral judgment for the latter is to put the former in charge” (Modrak, 213).

In the reality in which the person is functioning . . . it is hard for that person to animate herself as a person capable of practical syllogisms that are not subservient to the syllogisms of her dominator (2003, 58).

Though perhaps difficult for her to access, the woman does exist in “worlds” where she is not subservient—the power of arrogant perception is not all-pervasive. From these “worlds” she recognizes herself as one who is able to form liberatory practical syllogisms. In this respect, Aristotle’s concept of practical reason enriches Lugones’ articulation of what she comes to call ‘praxical thinking (Lugones, 2003).

Another curiosity arises within Lugones’ already unusual use of Aristotle, regarding citizen women’s reasoning and the situation of the slave. Lugones restates the situation of the slave on Aristotle’s account. The slave:

Cannot reason his own syllogism. The master reasons and the slave does. For that reason, it is difficult to imagine the slave after working hours except as an animal: resting, eating, roaming around, copulating. But these are not human acts when performed by the after-hours slave since they are not acts that are the end of practical syllogisms. . . (Lugones, 56).

How should we understand women’s lives under arrogant perception? Women are not slaves, but also are not as free as male citizens. Any reasoning done by the woman under arrogant perception reflects the arrogant perceivers’ interests:

Two practical syllogisms enter into actions performed by women who are successfully arrogantly perceived: hers and his. There is no denial that the woman can engage in practical reasoning. But because she can, the alternatives from which she can choose and thus her conception of her well-being at the moment of action must be altered, manipulated. She then ‘chooses’ between alternatives that she would not have chosen except for the arrogant perceiver’s mediation. Once the arrogant perceiver manipulates her alternatives, she proceeds to reason practically and she chooses the alternative the arrogant perceiver wants her to choose. So, the effect is the same as in the case of the slave (Lugones 56).

Lugones makes a curious statement when she claims that the ‘effect is the same’ for the woman and the slave. ‘Curious,’ because she risks eliding the difference between citizen women and slaves when the differences, as Aristotle frames them, are significant, both politically and epistemologically. Given dialogues about racism in the feminist community, including a justifiable caution to white women by women of color to refrain from equating gender

oppression with racial oppression, a bit of clarity is needed to draw out Lugones' meaning here.

Elizabeth Spelman, in accordance with the concern expressed by Lugones and many women of color about a racist eliding of difference, has pursued the significance of this distinction between citizen women and slaves of both sexes.¹ To explicate Lugones' point about women, slaves and arrogant perception, I will consider differences in virtue for women and for slaves and differences in their experience of arrogant perception and resistance to it. Rather than dismiss Aristotle's views of women and slaves as archaic (though influential up to the nineteenth century and beyond) or indefensible (which they are), I want to examine them in the context of arrogant perception. Aristotle is a definitive example of the arrogant perceiver; that is obvious. But an examination of how he goes wrong and what his perception fails to recognize can further explain the workings of arrogant perception in our own time.

The Status of Women and Female Slaves

The coalescence of gender and 'race' (see footnote 4) suggests that relations between groups of oppressed persons was complicated by this mix. Some sources suggest that citizen women were responsible for ordering, even ruling the household, and so were in positions of authority with respect to female and perhaps some male slaves (Freeland 1994, 154, 181n38). The citizen woman may have had some power to rule, but none that Aristotle admits. The female slave, on the other hand, is not recognized by Aristotle as a woman. What specifically was the experience of the female slave? It is likely that she faced increased vulnerability to sexual violence, by citizen men and male slaves, being denied the "protections"/proscriptions of the citizen woman. Furthermore, citizen women may not have been ready allies of female slaves, the contrary is more likely. One reason for Aristotle to distinguish between

¹Spelman argues that the status of slave is a mixture of class and race as we understand both in the modern sense. Neither term fully captures the distinction so she uses the term 'race' in quotation marks to highlight this point (Spelman 124, note 44). Spelman argues that feminists cannot simply consider citizen-women in determining the status of women in Aristotle's work; one must consider all persons designated female. She argues: "one's 'gender' identity is determined by sexual and 'race' identity . . . Slaves are without gender because, for Aristotle's purposes, their sex doesn't matter. In any world in which, for some people sex is made to matter—positively for males, negatively for females—then it also matters a lot if your sex doesn't matter" (Spelman 119). Gender differences in Aristotle are inextricably linked to ones "race." As Spelman puts it, to have a gender identity is to have a kind of "race identity" as well, specifically, a privileged race identity (118). Thus we do well to follow her direction in considering that, though women are subordinated to men in the polis, it is also the case "according to [Aristotle's] catalog of persons, functions, virtues and natures some men are naturally subordinate to some women (or in any event not naturally superior to those women); some women are not naturally subordinate to some men; and some women are naturally subordinate to some other women" (Spelman 110).

women and female slaves is to secure the birthright of male citizens. In the *Politics* he defines a citizen as one born of citizen parents (1275b22-23). Children of slaves or of mixed parentage are excluded from citizenship.

Aristotle denied gender distinctions to slaves, and thus denied male slaves the power to rule female slaves in the way men rule women, because rulership requires the ability to deliberate. Citizenship or the withholding of it establishes one kind of hierarchy; gender distinctions establish another, distinct, but interconnected hierarchy. Contemporary considerations of race and gender still often fail to consider the coalescence of these forms of oppression.

Virtue and Resistance under Oppression

Another effect of the gender distinction among citizens and the class distinction between citizens and slaves was to determine the kinds of virtues one was expected to cultivate. Citizen men cultivated the virtues of leaders such as courage and justice; citizen women cultivated subordinate virtues such as obedience, silence and a more modest form of courage; male and female slaves cultivated the virtue of adherence to duty (*Politics* 1260a14-1260b8). For those with deliberative reason, that is men or women, acting rightly means one is either virtuous or continent. Continence is choosing rightly, but against one's desires, whereas virtue is to desire rightly and of course choosing what one rightly desires (*NE* Book VII).

Women (and men) make choices in relation to their desires, either pursuing them or not. Desires shaped by arrogant perception may not be what they would have been otherwise. For example, Aristotle cautions against lasciviousness, a vice specific to women. What did Aristotle fear women's desires might be without this proscription? Perhaps women would desire men other than their husbands, perhaps other women, but these options are foreclosed by the definition of a virtuous woman. Another example: silence is a virtue for a woman (but not a man) (*Politics* 1260a30-1). A woman's *desire* to speak may exceed the norm, but she conforms in *action*. In such a case, her choice makes her continent, not virtuous.

Given the above analysis of arrogant perception and its effect on the practical reasoning of women, we could speculate: If a woman is virtuous, it could be that her desire has been largely determined by arrogant perception, since she now desires what she ought to desire as a woman. If a woman is continent, perhaps her desire has not been arrogated. Her actions, conforming to the dictates of virtue against her desires, could also be described as coerced. If women are making coerced choices when choosing under oppression, it isn't clear that they are making virtuous choices at all, in Aristotle's sense. If all women's action in conformity with virtue is coerced then we could say there are no truly virtuous women in the *polis*. Women are not virtuous, merely compliant, though Aristotle would not recognize this.

In any case, it is safe to claim that some action is coerced; for example, some women remain silent against their desires. These 'illicit' desires offer a

path to a “world” outside of arrogant perception. If a woman can maintain her own desires, whether or not they conform to convention, her desires might provide an avenue of resistance.¹ If women understand themselves as *continent* (or incontinent) rather than *virtuous*, they might recognize their (thwarted) desires as working in their own interests, and thus worthy of pursuit regardless of arrogant perception. To resist arrogant perception, women must reconstruct and reclaim their own desires and choose to act on them.

Regarding the virtues of slaves, on Aristotle’s view slaves cannot be incontinent since that requires the power of deliberation. Slaves can only be obedient or disobedient (1260a34-7). It is significant that Aristotle does not recognize disobedience as a choice, that is, disobedience is not acknowledged as a form of agency. According to Aristotle, the slave cannot willfully disobey. Disobedience might be explained by the slave’s being motivated by bodily desires that overrule the guidance of the master (*De Motu Animalium* 701a28-38)². In any case, on Aristotle’s construction, the slave does not deliberate; the slave merely acts, as Lugones points out in the passage cited above.

Yet, disobedience can be willful, and it can be a kind of advantage if the arrogant perceiver fails to read disobedience as willfulness. To use a modern example, Robin Kelley has done important work documenting the resistant actions of subordinated people who use the ignorance of the dominators. Their actions are marked as stupidity or laziness, which harms them, but also protects them because it masks their resistance.³ Such resistance could maintain a sense of a “world” in Lugones’ sense not completely dominated by arrogant perception. As conceptualized by Aristotle, virtue for those classified as slaves is about bodily compliance. There is a particular emotional and intellectual suffering that results from being constructed as not capable of deliberation. For those denied recognition of the power to deliberate, dismantling the structure of oppression is not a matter of acting on one’s own syllogisms rather

¹Frye, 1983b, 77.

²“For when a man actualizes himself in relation to his object . . . what he desires he does at once. For the actualizing of desire is a substitute for inquiry or reflection. I want to drink, says appetite; this is drink, says sense or imagination or mind: straightaway I drink. In this way living creatures are impelled to move and to act, and desire is the last or immediate cause of movement, and desire arises after perception or after imagination and conception. And things that desire to act now create and now act under the influence of appetite or impulse or desire or wish” (*De Motu Animalium* 701a28-38).

³“Studies of black North Carolina tobacco workers reveal a wide range of clandestine, yet collective strategies to control the pace of work or strike out against employers. When black female stemmers had trouble keeping up with the pace, black men responsible for supplying tobacco to the stemmers would pack the baskets more loosely than usual. When a worker was ill, particularly black women who operated the stemmer machines, other women would take up the slack rather than call attention to her condition, which could result in lost wages or dismissal. On the factory floor, where stemmers were generally not allowed to sit or talk to one another, it was not uncommon for women to break out in song. Singing in unison not only reinforced a sense of collective identity but the songs themselves—religious hymns, for the most part—ranged from veiled protests against the daily indignities of the factory to utopian visions of a life free of difficult wage work” (Kelley 1996, 18). Kelley’s account highlights the collective nature of such resistance.

than another's, initially it is the struggle to be recognized as having the power to form syllogisms.

Women of the citizen class and slaves of both sexes experience arrogant perception differently. But one could raise the objection that women under arrogant perception are reduced to the same rational ability as that assigned to slaves, meaning they function the way slaves function under arrogant perception. Both are, of course, capable of the kind of reasoning denied to them, but under arrogant perception, these abilities are blocked by coercion and manipulation. Certainly there are cases where a woman is so severely abused that she is reduced to a completely abject state. Kathleen Barry documents this process in detail.¹ In cases of female sexual slavery, the term 'slave' is not an overstatement. In other cases, the status of slave is the benchmark against which other forms of oppression are measured.

A woman subject to the arrogant eye is not reduced entirely to the level of slave—and, as Frye's analysis implies, it is precisely the difference that is so crucial. She must possess enough will to be said to 'choose' her condition. Again, as Frye points out, it is important to consider the context in which choice is structured. The distinction between 'slave' and 'woman' is crucial to understanding the nature of different forms of oppression because the understanding of women's oppression (as distinct from that of female or male slaves) means recognizing the role of will or choice as an element of the dynamic.

The citizen woman is perceived by Aristotle to be choosing to perform the actions willed by the arrogant perceiver. She may share this perception, making her appear to be 'complicit' in her subordination in a way that does not apply to slaves. On the other hand, the survival of slaves and those similarly subordinated necessitates maintaining the ignorance of the oppressor regarding their deliberative powers. They are not perceived as 'choosing subordination because they are not perceived as 'choosing' at all.

To return to Lugones' statement regarding women and slaves, she does not say citizen women are made into slaves by arrogant perception. She recognizes the different epistemological status of each in Aristotle's concept of practical reason. Lugones focuses on action: she says 'the effect is the same,' which is true. Both are coerced into obedience. What differs is the means of coercion.

Conclusion

In this paper I have considered the use María Lugones makes of Aristotle's practical syllogism in articulating the dynamic of oppression, specifically, what she and Marilyn Frye call arrogant perception. Lugones appropriates Aristotle's concept of practical reason by drawing parallels to Frye's theory of agency under oppression.

¹In *Female Sexual Slavery*. Her account of this process forms the basis of Frye's argument regarding arrogant perception.

Aristotle isn't your average arrogant perceiver. He doesn't simply want it his way; he works out a theory of practical reason that shows exactly how he should get his way. And in doing so, he shows us where arrogant perception gets it wrong. Furthermore, arrogant perception is not limited to relations between men and women. Frye's concept of 'whiteness' articulates this type of perception on the part of white people with respect to race (1992), and Charles Mills (1997) develops a book-length analysis of the 'epistemology of ignorance.' Though we reject Aristotle's archaic concepts of women and slaves, we can benefit from studying them. We can learn something about how contemporary gender, race and class relations still carry a bit of Aristotle's ghost in the background.

Works Cited

- Aristotle. (1912 [1949, 1958]). *De Motu Animalium*. Translated by A.S. L. Farquharson. In *The Works of Aristotle* vol. V, edited by W.D. Ross. Oxford: Clarendon Press.
- _____. (1983 [1986]). *Nicomachean Ethics*. In *Aristotle: Selected Works*. Translated by Hippocrates G. Apostle and Lloyd P. Gerson. Grinnell, Iowa: The Peripatetic Press.
- _____. (1921 [1946, 1952, 1961]). *Politics [Politica]*. Translated by Benjamin Jowett. In *The Works of Aristotle* vol. X, edited by W.D. Ross. Oxford: Clarendon Press.
- Barry, Kathleen. (1979). *Female sexual slavery*. Englewood Cliffs, New Jersey: Prentice-Hall. Cited in Marilyn Frye 1983b, *The Politics of Reality: essays in feminist theory*. Freedom, California: The Crossing Press.
- Cooper, John M. (1984). *Reason and human good in Aristotle*. Cambridge: Harvard University Press.
- Freeland, Cynthia. (1994). 'Nourishing speculation: A feminist reading of Aristotelian Science.' In *Engendering origins: critical feminist readings in Plato and Aristotle*, ed. Bat-Ami Bar On. Albany, New York: State University of New York Press.
- Frye, Marilyn. (1983a). 'Oppression.' In *The Politics of reality: essays in feminist theory*. Freedom, California: The Crossing Press.
- _____. (1983b). 'In and out of harm's way: arrogance and love.' In *The Politics of reality: essays in feminist theory*. Freedom, California: The Crossing Press.
- _____. (1992) 'White Woman Feminist,' in *Willful virgin: essays in feminism, 1976-1992*, Freedom, CA: Crossing Press.
- Kelley, Robin D. G. (1996). *Race Rebels: Culture, Politics, and the Black Working Class*. NY: Free Press.
- Lugones, María. (2003a). 'Structure, anti-structure, and agency under oppression.' In *Pilgrimages/peregrinajes theorizing coalition against multiple oppressions*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc.
- _____. (2003b). 'Playfulness, "world"-traveling, and loving perception.' In *Pilgrimages/peregrinajes theorizing coalition against multiple oppressions*. Lanham, Maryland: Rowman & Littlefield Publishers, Inc.
- Mills, Charles. (1997). *The Racial Contract*. Ithaca: Cornell University Press.
- Modrak, Deborah K.W. (1994). 'Aristotle: Women, deliberation, and nature.' In *Engendering origins: critical feminist readings in Plato and Aristotle*, ed. Bat-Ami Bar On. Albany, New York: State University of New York Press.

An Anthology of Philosophical Studies

Spelman, Elizabeth V. (1994). 'Who's who in the polis?' In *Engendering origins: critical feminist readings in Plato and Aristotle*, ed. Bat-Ami Bar On. Albany, New York: State University of New York Press.

CHAPTER FIVE

The Examined Life: Outline of a Neo-Socratic Argument

Joe Mintoff

Socrates famously proclaimed in his *Apology* that the unexamined life is not worth living. More precisely, he claimed that 'to let no day pass without discussing goodness and all the other subjects about which you hear me talking and examining both myself and others is really the very best thing that a man can do, and that life without this sort of examination is not worth living' (Apl 38a).¹

This is not very plausible, exactly as it stands. For consider the average person's life, one without Socratic cross-examination. It involves instrumental activities (eg, earning an income), which preclude engagement in philosophy; and non-instrumental activities (eg socializing with friends), which either do not raise the issue of how to live, or, if they do, almost never involve the sort of relentless cross-examination Socrates inflicts upon his interlocutors. Isn't it just hyperbole to say that such lives are not worth living? (Schmitt 2004: 310). If we really think that certain lives are not worth living, then we mean very bad lives indeed: like those filled with constant pain and suffering from which there is no relief, or else those heroically sustained by respirators and tubes snaking into often unconscious bodies. If the much-quoted claim in the *Apology* is to be more than just hyperbole, then the same must apply to it (Schmitt 2004: 310). But clearly the same does *not* apply: we have no dark thoughts about the average person's life, which we can see is very far removed from the sort of suffering described above, and which we think would still be worthwhile even if it could never be enriched by Socratic cross-examination. We cannot agree that a life without Socratic examination—that is, the average person's life—is literally not worth living.

Nevertheless, some of Socrates' discussions point to a more plausible claim concerning the relationship between philosophy and a well-lived life. The chief claim in this paper is that if one does aim to live well (as most people do) then one ought to do philosophy, and its chief task is very briefly to explain how to defend Socrates' claim, and in the process to explain what it amounts to. We finish by considering an important objection.

¹Abbreviations and sources for the Platonic dialogues are as follows: Euthyphro (Eup), Crito (Crt), Apology (Apl), and Phaedo (Phd) from Plato 1993, *Last Days of Socrates*; Ion (Ion), Laches (Lch), Lysis (Lys), Charmides (Chr), Hippias Major (HMa), Hippias Minor (HMi), and Euthydemus (Eud) from Plato 1987, *Early Socratic Dialogues*; Alcibiades I (AlcI) from Plato 1997, *Complete Works*; and Gorgias (Grg), Meno (Men), Protagoras (Prt), Republic (Rep), Symposium (Smp) as indicated in the bibliography.

In the *Euthydemus*, Socrates exhorts Cleinias to take up philosophy. He caps his exhortation with a neat summary of his argument: '[t]he desire for happiness is universal, and we found that happiness stems from use – correct use – of things, and that correctness, in turn, and good luck, are products of knowledge; it apparently follows that everyone should be expending all their efforts on making themselves as wise as possible' (Eud 282a), and so, assuming wisdom is teachable, that philosophy is essential (Eud 282d). This argument has four steps.

(1) Socrates starts his exhortation to Cleinias with the premise that everyone wants success (Eud 278e) or *happiness* (Eud 282a).

But *is* this what everyone really wants? Elizabeth Telfer points out, for example, that some people seem to sacrifice a good life to *some strong urge*: 'a man might quite knowingly and deliberately sacrifice his happiness to pursue revenge' (Telfer 1980: 23). Others sacrifice it to *duty*: 'a man may think he can achieve happiness only by doing something he considers wrong, and so he may renounce happiness ... [or] a man may choose to sacrifice his life at the call of duty' (Telfer 1980: 23), and he may be of such a deontological cast of mind that he is indifferent or hostile to thinking about his life in terms of whether it realizes some desire for happiness.

Rather, perhaps what everyone *should* want is happiness. Talking to Callicles on another occasion, Socrates claims we can all agree that 'the good [or happiness] is the end of all actions, and that for the sake of it we should do all other things, not do it for the sake of other things' (Grg 499e). He does not say why, but perhaps the idea is that the prudent approach to one's affairs is to order one's life as a whole, and that anyone who does so will thereby do all things for the sake of single (though perhaps compound) end, in which they take their happiness to consist (cf. MacDonald 1991: esp pp 46-59, and Becker 1998: ch 6 esp pp 114-118).

We all have everyday aims we take seriously, which we are committed to judging have *some* value. We do not pursue each everyday aim in isolation and without foresight, but show *some* sensitivity to, and on *some* occasions explicitly deliberate about, higher aims (if any), conflicting aims (if any), and later times, since otherwise we risk a poor realization of all those aims. When our everyday aims conflict (as they often do), we must and do decide *in outline and subject to revision* what we really want for our day and how much we really want it, which decision guides our subsequent actions and commits us to a conception of what is better or worse for that day as a whole. It would be folly not to consider our everyday aims, not to decide which of them are most important, and not to aim to well-realize them all. But life is a day writ large, and—this is the central point—the same things that make it reasonable for a person to strive for coherence among their everyday aims will also make it reasonable for them to strive for coherence among their aims at the highest level. We seriously pursue not only everyday aims, but (on pain of regress) we also have *highest* aims, which we also take seriously, and which we are committed to judging of highest and greatest value. And so, for exactly the same reasons, we are committed not to ignore the fact that each day is part of a

life more broadly, and show *some* sensitivity to, and on *some* occasions explicitly reflect about, our highest aims, as far into the future as they cover. And since our highest aims are likely to conflict, then, for exactly the same reasons, we are logically committed to decide *in outline and subject to revision* what we really want for our *life* and how much we really want it—that is, to decide upon an ultimate end—which subsequently guides all we do during our life and which commits us to a conception of what is better or worse for our lives as a whole. When so much more is at stake, it would be even greater folly not to consider our highest aims, not to decide which are most important, and not to aim to well-realize them all.

(2) Continuing, Socrates argues that *wisdom* is the only good. Happiness involves the possession of good things, and presumably this means financial goods (eg, wealth), bodily goods (eg, health), social goods (eg, power), moral goods (eg, courage, wisdom), and good luck to boot. It turns out this list requires significant revision. First, *good luck* is just wisdom: in any area of expertise, one can expect to have better luck if one relies on an expert rather than the non-expert, and so, Socrates concludes, wisdom causes luck and never fails in its goal (Eud 279d). Further, the possession of good things will lead to happiness if and only if they are correctly used (Eud 280b-e), and this will be so if and only if they are accompanied by wisdom. All this means that, in and of themselves, the supposed "goods" such as wealth, health, power, and courage depend for their "value" on the presence of wisdom, and so, Socrates concludes, 'nothing else is good or bad, and that of the pair we've been discussing, wisdom is good, ignorance bad' (Eud 281e).

This section of the argument is far from convincing. Happiness is not always the correct *use* of possessions, since there are some good things of which the mere possession (a good reputation, perhaps), or the mere use without "correct" use (autonomy, perhaps), is at least a part of happiness (Waterfield 1987: 306). Further, Socrates fails to establish that correctness in the use of one's possessions is *always* the product of knowledge: he argues that wisdom causes good luck, since it is more likely to be successful than ignorance (Eud 279d), but sometimes it isn't (eg when there are no goods to use), and even when wisdom is more likely to be successful, it doesn't follow that it *invariably* leads to success (Santas 1993: 45-7). Further still, even if happiness is the correct use of possessions, and is assured by wisdom, it simply does not follow that wisdom is the only good, since all this would establish is that wisdom is an instrumental good, and thus that there must be some *other* good – happiness – for which it is the instrument.

Even so, we can still say that living well *requires* wisdom. But there are different grades of wisdom, and living well does not require them all. (i) One's **success** in living is proportional to the extent that does well, that is, to the extent that one does what happens to be the right thing, and, if applicable, for the right reasons, in the right way, and at the right times (cf AlcI 107d)—on the assumption, of course, that there is often a difference between doing the right thing or not. (ii) One **knows how** to live well when one has the reliable ability to do the right thing, which will involve the ability to choose the right final

ends and the ability to select the right means of realizing them. Know-how is clearly different from success: poor painters sometimes produce good paintings by luck; and good painters bad ones (though we rarely get to see them); and, if an important aspect of living well is the value of the outcomes one realizes, then a similar point will apply to living well. (iii) Someone is a **teacher** of good-living when they are able to teach another how to live well, which will presumably include telling them things, but also (eg) giving them the right setting in which to learn, showing them things to imitate, getting them to practice, and so on. Teachers and practitioners are not necessarily identical: some good painters are poor instructors, and possibly vice versa; someone with the ability to make the right decisions in their own lives may be unable to pass on that know-how to others. Finally, (iv) someone is a **sage** when they have the ability to correctly say sufficiently general or systematic things about what it is to live well, the very height of which is **theoretical knowledge**, the ability to provide a correct general and systematic account of these matters. Sages are not identical to practitioners: while the latter may be able to judge their own performance, it does not necessarily follow that they can judge that of others, or say anything interestingly general or systematic about it. Nor are sages identical to teachers: good painting-critics might be able, post hoc, to judge and theorize, but that does not imply any ability to give advice concerning an ongoing painting or to train people to paint; conversely, while good painting-teachers will presumably be good judges of their students' work, they need not have an overall theory of painting; and the same applies even more to an activity as complicated as living.

The relation between living and these forms of wisdom seems to be as follows. (i) Living well *entails* success in living—one will simply not count as living well if one rarely does the right thing, or, if appropriate, rarely does so for the right reasons, etc. (ii) Living successfully does not entail but *probably requires* one to know how to live well—Socrates was right that wisdom is more likely to be successful than ignorance, but wrong to think that this entails that it will be successful. This means that living successfully requires knowing the function of the goods in one's life in sufficient depth to be able to make the right choices in relation to them, but, as we are about to see, this need not imply any great ability engage in extensive technical discussion of the matter. (iii) In turn, knowing how to live well *probably requires* learning how to live well from some teacher, or (if that fails) learning how to live for oneself using some reliable method—knowing how to do *anything* well requires being taught, or learning for oneself. And nor is *living* well an exception to this, as some think who would find it demeaning to take lessons in good living. For there are modes of instruction which respect—nay, develop—the student's ability to think for themselves in the relevant area. And, more provocatively, there is in any case no harm approaching the supremely wise on any issue—for if it is more important for us to decide that issue autonomously, then they will know it (lest they not be supremely wise), and they will therefore tell us to decide for ourselves (assuming they are prepared to give sincere advice); and if it is more important for us to get the right answer on that issue than to decide

autonomously, then the supremely wise will tell us to let *them* decide, and there can be no objection to our letting them do so if it really is so important that we get the right answer. However, (iv) *pace* Socrates, living well does not necessarily require sagacity—for example, Australian aborigines know how to identify edible plants, and know how to pass that skill onto their children, even though they lack the modern nutritionist's theory-based understanding of diet. There is no reason to suppose that living is any different.

(3) Having identified happiness with wisdom, Socrates claims that we should be expending all our efforts on becoming as wise as possible (Eud 282a). Clearly, he must be assuming that *no one already has the wisdom* sufficient for happiness, but again he does not say why to his interlocutor, Cleinias.

He does offer reasons in other dialogues. He does not think that the many (ie, the general public) possess the requisite wisdom (eg, AlcI 111a ff). In response to just such a claim by Socrates, Alcibiades claimed that the many are teachers, for example of Greek, thus implying that they could also teach him (and so possess wisdom) about justice (cf. Lch 184c, Crt 47c). However, as Socrates points out, there is an important different between the two: the many (taken collectively) agree about what is the correct way of speaking, but they notoriously disagree about justice (and, we might add, about happiness), and no one can be a teacher who contradicts themselves. Nor does Socrates think that reputedly-wise individuals have this wisdom (eg, Apl 21b). He reports at his trial how he examined politicians, poets, and skilled craftsmen, but found that all lacked wisdom, and worse still lacked the knowledge that they were in this deplorable state.

Socrates is on pretty strong ground here. We need reflect only on our own case. First, we all make mistakes, and know we do. Second, our general decision-making habits often involve unreliable ways of thinking: the errors we regularly make in reasoning are so well documented that they have names (see Weiten 1998: 332-5). For example, in estimating probabilities, people typically employ the so-called representativeness heuristic (judging the probability of B given A by evaluating the degree to which A resembles the stereotypical B), which is quick but not always reliable. Third, the kind of life we currently lead almost certainly involves inconsistencies: if we were asked to explain why our chosen ends are right, and if our answers were cross-examined, then in all likelihood we would be reduced to contradiction and perplexity. We who are acquainted with philosophy know how it can sometimes lead us from things we believe, by inferences we accept, to other things we certainly do not believe, and this would presumably also apply to those innocent of philosophy. Now, if we suddenly discovered that our *doctor* sometimes makes mistakes, that he regularly takes decisions in ways known to be unreliable, and that he couldn't explain his prescriptions to a layperson without contradicting himself, then we would without a moment's hesitation seriously doubt whether he knew how to improve our health. But, if our ultimate aim really is to live well, then living well is more important than living healthily (Grg 511b, 512d, Crt 48b). Hence,

just like the doctor, we should seriously doubt whether we know how to live as well as possible.

(4) Socrates finally concludes that philosophy – the love of wisdom – is essential, and he exhorts Cleinias to take it up (Eud 282a).

One approach to philosophy is suggested by what Socrates says next, that one should become a deferential student of some wise teacher (Eud 282b). However, while this might be a sensible policy were it easy to recognize any such teacher, we have already seen that Socrates thinks that no one actually has the requisite wisdom, and we have seen why we might doubt our own ability to immediately recognize them even if they existed.

A different approach is suggested by what Socrates himself does. During his discussion with Callicles, Socrates makes the following, much remarked upon, claim: 'These things which appear true to us earlier in the previous arguments are held firm and bound down, so I say – even if it is a bit impolite to say so – by iron and adamantine arguments; so at least it appears so far. And if you, or someone more vigorous than you, doesn't untie them, no one who says anything besides what I say now can be right' (Grg 508e). The suggestion here is that the way to justified beliefs about happiness—Callicles and Socrates are talking, in particular, about whether the happy man is just—is to subject one's views to the sort of cross-examination Socrates constantly inflicts on his interlocutors. There are two ideas here.

The first part of Socrates' suggestion is that our sincere beliefs about the wisdom of given aspects of our lives must be examined, and that they must be examined *adequately*. Clearly, sitting only a brief test conducted by a poor examiner hardly displays openness to cross-examination, and (even if we passed) would be insufficient to remove our doubts. However, it *would* be sufficient if our beliefs about a given aspect of our lives pass all the important tests at the hands of skilled examiners. Now, even though the distinctive skills of academic moral philosophers do not make their moral prescriptions more likely *right* (do they not disagree as much amongst themselves as the many, whom we dismissed above as teachers of wisdom?), it is plausible claim that these skills do make their prescriptions more likely *reasonable*, and so the skilled examiners we are after are academic moral philosophers, and the important tests will simply be the ones they would say are important. We need to be more specific.

For any specific issue on how to live, there will be certain people who would be more skillful examiners on that issue than any others, and, for any specific view on that issue, there will be certain tests which these people would agree it must pass if it is to be justifiably believed. These ideal examiners are those who are best able to find true refutations of people's real beliefs. This ability is twofold. First, there is the ability to *inspect* people's assertions: the psychological insight into others to be able to prompt clear and concise statements from them, understanding of their points of view, and the skill to ask questions which minimize the risk that they misspeak. Second, there is the ability to *test* people's assertions: the logical acumen to uncover objections to others' claims, using premises and inferences they would accept. Socrates is

often thought to be just such an examiner, but he is no longer with us, so we must make do with the next best thing—academic philosophers. Consider, then, some issue directly concerning some aspect of how people are living, the sort prompting everyday deliberation or critical examination—thus not in the first instance one relating to (eg) the nuances of rule utilitarianism, but rather one concerning (eg) the balance between friendship and morality when the two conflict. Let us say that the **key academic literature** on that issue consists of the widely-reviewed books produced by publishers highly-regarded by academic philosophers, the much-cited articles appearing in prestigious journals and anthologies, and the works discussed at some length in widely-read survey articles and widely-prescribed introductory texts, and so on. This literature sets the final *practical* standard for cross-examination. For any specific issue on how to live, there will be certain flesh-and-blood *academics* who are more skillful discussants of that issue than any others: namely, those who know the key academic literature and have the requisite critical skills (evidenced by publications), and are more skillful at inspecting people's assertions than the others (evidenced by teaching?). For any specific view on that issue, there will be certain *tests* which these academics would agree it must pass to be justifiably believed: these are the objections typically appearing in the key academic literature, but also any objections that these academics *would* come up with *were* some novel view not already appearing in the literature to be proposed. *For practical purposes*, the key literature determines who the key examiners are, and the questions these examiners would ask determines what the key tests are.

Passing these tests of these examiners, therefore, implies that our ethical beliefs are justified. Let us say that our thinking on some issue has received **sufficient academic testing** when it has been subjected to all the important tests at the hands of skilled examiners (as determined by the key academic literature), and perhaps other testing by other examiners (as likely follows from an openness to cross-examination). We can say that *if our thinking on some issue (large or small) has received sufficient academic testing, then we are justified in relying on that thinking as a whole if and only if it has passed every relevant test*. This applies to our thinking about how to live: before we have sufficiently tested our living as a whole, we should seriously doubt how well we are living; afterwards—not that we are ever likely to complete so large a task—we can be confident that we are living well if and only if our lives have passed every test. And it applies to our thinking about some specific aspect of our lives: before we have sufficiently tested (eg) some ethical belief, we may or may not be justified in believing it; afterwards, we are justified in believing it when and only when it has passed every test.

But there is no point sitting these exams if we cannot pass, and so—this is the second part of Socrates' suggestion—we are further required to become students of ethical philosophy. We must seek instruction in philosophy to become inquirers into the art of living and then regularly reflect (better with the help of others) to ensure we act on defensible ideas about how to live. If we do,

we will thereby have acquired the ability to uncover for ourselves how to live better. Again, we can be more specific.

We should take what we might call the **Maieutic approach**, modeled in part on the relationship between an academic philosopher and a group of three or four graduate students. The ultimate aim of this approach is to enable people to think for themselves about how best to live, while being answerable to the relevant academic literature, and the primary means to this end is an academic mentor. The only knowledge presumed at the start is reading ability and verbal communication skills.

In the early stages, the pedagogical aim is simply to create a so-called community of inquiry, with the mentor functioning very like a teacher. He attempts to inculcate the skills students will need to generate justified beliefs about how to live. He will encourage *openness* in students (eg, a preparedness to acknowledge different perspectives and viewpoints). He will promote their ability to *express* their own views clearly and concisely, making sure they understand the work-a-day concepts (eg, means vs ends) they need in order to do so. He will instruct them in *critical reasoning* (eg, how to spot and diagnose faulty reasoning). But more than this—and much harder than this—he will also nurture their capacity for *creative inquiry* (eg, their ability to ask fruitful questions, to explore alternatives, and so on). And, through all this, he will also encourage the intellectual *virtues* they will need if their criticism and inquiry is to be fruitful (eg, the courage and perseverance to keep trying to resolve their philosophical dilemmas even though answers do not appear right away).

In the later stages, the pedagogic aim becomes to ensure that the new community of inquiry is answerable to the relevant philosophical literature, with the mentor now functioning as a catalyst, by being a living embodiment of that literature. The content is determined entirely by the students: in particular, by the specific questions prompted by their chosen life-ends. The learning materials are mentor-selected academic publications relevant to the student-chosen philosophical content (which the students may or may not be required to read). The teaching method is centered on the mentor, who needs to be a skilled examiner and good intellectual midwife. As a good examiner, they take any assertion by their students and skillfully subject them to the main problems needing to be resolved if it is to be sensibly believed (as determined by the academic literature). Initially, the students are not likely to pass. But, unlike some examiners, our mentor is also a skilled facilitator. They help their students pass: they help them develop their own thoughts in response to the problems to which they have been subjected, by carefully helping them uncover what they are already inclined to think, and gently suggesting how these nascent ideas might be developed (as suggested by the academic literature). And the assessment is explicitly Socratic: can the student withstand cross-examination conducted by their mentor? So long as it is done *cautiously*, this instruction and facilitation is consistent with students' uncovering for themselves how to live better.

We may go so far as to claim, then, that *to receive instruction in the art of philosophy, and then to let no important opportunity pass without discussing*

what it is to live well and all related subjects, by examining both oneself and others, is *required by* the very best thing that one can do (namely, to live well), and that living without this sort of examination *runs a serious risk of one's life being wasted*. This is not hyperbole. If our ultimate aim really is to live well, then we will believe that the most important thing is to do so, but wasting our life implies living poorly (wasting anything implies using it poorly), and so we will also believe that it is *most important* not to waste our life.

References

- Becker, L. (1998). *A New Stoicism*. Princeton, NJ: Princeton University Press.
- MacDonald, S. (1991). 'Ultimate Ends in Practical Reasoning: Aquinas's Aristotelian Moral Psychology and Anscombe's Fallacy.' *Philosophical Review* 100: 31-66.
- Plato. (1956). *Protagoras and Meno*. Harmondsworth: Penguin, translated by W Guthrie.
- Plato. (1960). *Gorgias*. Harmondsworth: Penguin, translated by W Hamilton.
- Plato. (1974). *Republic*. Harmondsworth: Penguin, translated by D Lee.
- Plato. (1987). *Early Socratic Dialogues*. Harmondsworth: Penguin, edited by T J Saunders, translations by T J Saunders (Ion), I Lane (Laches), Watt (Lysis, Charmides), R Waterfield (Hippias Major, Hippias Minor, Euthydemus).
- Plato. (1993). *Last Days of Socrates*. Harmondsworth: Penguin, edited by H Tarrant, translated by H Tredennick.
- Plato. (1994). *Symposium*. Oxford: Oxford University Press, translated by R Waterfield.
- Plato. (1997). *Complete Works*. Indianapolis: Hackett, edited by J M Cooper.
- Santas, G. (1993). 'Socratic Goods and Socratic Happiness.' *Apeiron* 26: 37-52.
- Schmitt, R. (2004). 'Is the Unexamined Life Not Worth Living?' *Teaching Philosophy* 27: 307-319.
- Telfer, E. (1980). *Happiness*. London: Macmillan.
- Waterfield, R. (1987). 'Introduction to *Euthydemus*.' In Plato 1987: 297-311.
- Weiten, W. (1998). *Psychology: Themes and Variations*. 4th ed. Pacific Grove: Brooks/Cole.

CHAPTER SIX

The Milindapanha ('Questions of King Milinda'): An Early Encounter between Greek and Buddhist Philosophy

Jere O'Neill Surber

The *Milindapanha* (often translated as 'Questions of King Milinda') is the earliest documented encounter between Greek and Buddhist philosophical views. (See Bibliographic Note at the end of this essay for information on texts and citations.) It is regarded as a 'quasi-canonical' work of the Pali canon and seems to have enjoyed considerable popularity in some earlier Buddhist circles. Its dates are a bit uncertain, especially since several sections seem to have been added over the course of two or three centuries. The original core is usually regarded as the first three books, which date from about the 1st century BCE, while its later additions may extend as late as the 2nd century CE.

The work presents a sort of dialogue between a Buddhist monk and scholar named Nagasena and the Greek Bactrian king Milinda (his Pali name) or Menander (in Greek). There is no corroborating evidence to suggest that Nagasena was an actual historical figure. The first book of the work, which contains a sort of 'spiritual biography' of the monk beginning with earlier rebirths, suggests that he was well-educated, especially in the study of the *Abhidamma* (the scholastic compendium of Gotama Buddha's teachings in the *sutras* and a division of the 'Pali canon'), had become an *Arahant* (i.e. achieved release or enlightenment), and was serving as a revered teacher and leader of a sizeable group of monks at the time the dialogue occurs. Milinda/Menander, however, is relatively well documented as one of the Bactrian kings who seemed to have ruled a fairly extensive area (although there was more than one king of this name, so there has been some discussion about which the Menander of the dialogue actually was). One intriguing piece of evidence is provided by coinage of the time, which depicts a portrait of Menander together with recognizably Buddhist symbols, something that might be taken as confirmation, along with other documentary evidence, of the king's 'conversion' to Buddhism as related in the dialogue.

While the *Milindapanha* has received significant scholarly attention on the part of Buddhist exegetes and some modern philologists, it has rarely been discussed by philosophers conversant with early Greek and Hellenistic philosophy. As a result, although efforts have been made to determine Nagasena's specific viewpoint and place within Buddhist thought in relation to the various Buddhist sects of the time, little attention has been given to the

distinctively Greek ideas and assumptions that Milinda/Menander brings to the discussion. This results in a missed opportunity to engage what is clearly an early and significant document for what has later become called ‘comparative philosophy.’ This essay will be a tentative and exploratory attempt to probe what distinctively Greek elements might have been in play in the early encounters between Greek philosophy and Buddhist thought. Its main sections will correspond to specifically Greek ideas and themes discussed in the dialogue.

The Form and Style of the *Milindapanha*

If one begins with Greek models of the ‘dialogue’ form (especially those of Plato), then he or she might well have some reservations about calling the *Milindapanha* a ‘dialogue’ at all. True, it does have a dramatic setting, complete with a prologue somewhat resembling those occasionally found in Plato’s works. Likewise, it involves two main interlocutors who engage in philosophical exchanges. However, it seems to lack the sort of topical openness, fluidity of discussion, and occasional surprising dramatic twists characteristic of many of Plato’s dialogues. Rather, the discussion unfolds, by prior agreement of the interlocutors, as a series of ‘question-and-answer sessions’ about various points of Buddhist teaching, with the king asking the questions and the monk responding with the doctrinally ‘correct’ view.

The rather formulaic style of the discussion is reinforced by some specific features of the setting. In the prologue, King Milinda is portrayed as a monarch, one of whose passions is ‘posing questions’ to scholars. Prior to his meeting with Nagasena, we are told that he had already questioned many others, especially Brahmin thinkers, and had found their answers unsatisfactory. Also, he had attempted to question another Buddhist sage, but was disappointed since, from the start, the monk responded only with silence. When the king hears of the renowned monk-scholar Nagasena, who is prepared to ‘mow down his questions,’ the king resolves to seek him out and engage him in discussion. What is most interesting, however, is that the king’s questions tend to follow quite closely the structure of *abhidhamma* teaching, which, we have already been informed, is a specialty of Nagasena. Even allowing for a certain amount of ‘leading’ in the monk’s responses, the reader can only conclude that Milinda was already either acquainted with the *Abhidhamma* or content to allow his interlocutor to dictate the overall framework of the discussion. Whichever one concludes, the result is that the structure of the discussion, at least in the earliest parts of the work, is itself much more like a Buddhist ‘dhamma teaching’ than a Greek dialogue.

This (among other things) has led some scholars to speculate about the authorship of the work and its possible purpose. Although the *Abhidamma* and other commentaries upon it or the *sutras* are completely expository and lack any dialogal element, the dialogue form was well known to any student of the *sutras* of the Pali canon, since it is a device frequently employed in Gotama

Buddha's own teachings. The novelty of the *Milindapanha*, perhaps, consists in utilizing a form, familiar from the *sutras*, to provide an exposition or commentary on the *Abhidhamma* (and, in later parts of the work, other philosophical questions arising from it). The usual conclusion drawn is that the author of the earliest sections of the text was a Buddhist, possibly a member of a Sarvastivadin-influenced school, who had had some contact with, though no very sophisticated knowledge of, Greek philosophy (something entirely plausible given the long-standing interactions between the two cultures). On the other hand, as noted above and for the same reason, the king himself likely knew a good deal more about Buddhist thought than some interloper new to the region.

It would be a mistake, however, to assume that the king is a mere dramatic device designed to serve as a simple foil for the monk's exposition of the *dhamma* based upon *abhidhamma* doctrine. For one thing, the prologue and other episodes in the text clearly attempt to capture something of the king's personality and character. On this score, he is no more a dramatic cipher than some of Socrates' interlocutors in the Platonic dialogues. Further, although *abhidhamma* structure and style provide much of the framework for the monk's teaching, several seemingly sincere intellectual concerns and perhaps reservations of the king surface within the general structure deployed in the monk's responses. In what follows, I will note and comment on some of the most prominent, suggesting, in the process, that they do, in fact, represent points where an outlook influenced by Greek philosophy would encounter most difficulty with early Buddhist thought. Certainly the text overall inclines in a Buddhist direction, which is reinforced both by the king's very frequent formulaic response of "You are dexterous, revered Nagasena" at the end of many sections and the fact that the king is eventually 'converted' to Buddhism and ultimately dons the robes of a lay follower. But the traces of actual Greek philosophical ideas and concerns (and not just some Buddhist author's vague or distorted notions of them) are unmistakable.

Greek and Buddhist 'Ontology'

The initial meeting of Milinda and Nagasena begins in a rather odd but, in the end, quite penetrating way. In fact, the first move is initiated by the monk and goes immediately to the heart of what will prove to distinguish their views on many other topics as well. Indeed, it reveals that the monk is already aware of what will be the major obstacle in explaining the Buddhist viewpoint to the Bactrian king. When the king asks the monk his name, the monk replies that his parents and his fellow monks call him 'Nagasena,' but immediately adds, 'this "Nagasena" is only a designation, a label, a concept, an expression, a mere name because there is no person as such that is found.' This understandably puzzles the king, who responds with a whole series of questions which will later be treated in more detail in the ensuing discussion.

In fact, some of the questions the king poses to the monk (for example, whether ‘Nagasena’ is identical to any of his bodily parts, their ensemble, his mental states, their combination, etc.) is taken directly from the *sutras* and *abhidhamma* teachings. The king concludes, on the basis of Nagasena’s consistently negative responses to all of these that the monk has lied to him: ‘You, reverend sir, have spoken a falsehood, an untruth. There is no Nagasena.’

Nagasena responds by pointing to the chariot on which the king has come and asks that the king ‘show me the chariot.’ The predictable outcome is that none of the parts of the chariot are the chariot itself, but neither is there something apart from the parts or their assemblage that the word ‘chariot’ designates. So, the monk concludes, the king is equally guilty of misrepresentation.

Had he been a good Platonist, the king might well have invoked the ‘idea of chariot’ as that to which the word refers, but, in a more Aristotelian vein, perhaps invoking something like the ‘material cause,’ the king responds that it is ‘because of’ the various parts of the chariot ‘that “chariot” exists as a mere designation.’ Notice, however, that, by this time, the king seems to have already agreed with the monk that ‘chariot’ exists only as a ‘mere designation,’ albeit one that has a causal basis in its material parts.

Nagasena indicates that it is exactly the same with respect to persons and concludes with two very important points, seemingly beyond anything the king has himself considered. First, the monk states that ‘in the ultimate sense there is no person as such that is found,’ thus introducing the distinction that later Buddhist philosophy will refer to as ‘the two truths,’ that is, the difference between ‘conventional’ and ‘ultimate truth.’ Second, he quotes a verse from the *sutras*:

Just as when the parts are rightly set
The word ‘chariot’ is spoken,
So when there are the aggregates
It is the convention to say ‘a being.’

The king registers his approval of Nagasena’s explanation, at least to the degree that he has understood it. As the ensuing discussion reveals, however, while the king may have intellectually grasped the monk’s point, he remains as yet unaware of the full implications of the monk’s teaching of ‘non-self.’ Old habits are hard to break.

At stake in this opening exchange is nothing less than two diametrically opposed ‘ontologies.’ From the time of Parmenides and Pythagoras, as ramified in Plato’s theory of ideas or forms, and finally with Aristotle’s ‘substance,’ Greek thought was anchored by the conviction that ‘true being’ must be stable, unchanging, and permanent, while that which is variable and impermanent constitutes the realm of ‘appearance,’ ‘non-being,’ and ‘falsity.’ Of course, always co-present in Greek thought was an opposed tradition, stemming from Heraclitus and others such as Empedocles and advocated

during later times by such figures as Protagoras, which held that the only constant was change or flux. But the overarching trajectory of Greek philosophical thought remained the attempt to accommodate impermanence and change within an ontological framework that privileged permanence and stability. Nagasena, by contrast, represents a tradition that proceeded in exactly the opposite direction, since one of the Buddha's fundamental teachings was that all things are impermanent. This produced a counter-trajectory of philosophical reflection aimed, at least in part, at explaining how that which appears to be permanent and unchanging can be understood as resulting from a coming together or concatenation of processes. Put in slightly different terms, while the Greeks regarded 'appearances' as unreliable and deceptive and sought a stable ground on the basis of which they could be understood and explained, Buddhists regarded anything conceived as stable, fixed, or substantial as 'appearance' (and hence delusion) and sought to reveal how they were always the result of further processes (ultimately leading to the major Buddhist doctrine of *kamma*). Nagasena's task in attempting to 'convert' his Greek interlocutor to his own viewpoint was, therefore, formidable, in that it involved nothing less than a wholesale alteration of the king's most basic orientation toward 'being' and 'appearance' themselves. It is, therefore, no surprise that, although the king often registers his approval of the monk's explanations on particular points, his questions indicate that he continues to struggle with this fundamental difference in orientation between them.

Moral Responsibility

Probably the single implication of Nagasena's teaching of 'non-self' that most concerns Milinda is that, from a Greek perspective, if there is no permanent 'person' or 'self,' if the 'self' is not in some way an enduring substance or entity, then there are no grounds for holding individuals responsible for their actions or maintaining a distinction between a virtuous and vicious character. The king brings this up at several crucial points early in the discussion, indicating that, even though he may have grasped the monk's argument intellectually, his concerns remain acute. For instance, the king's initial response to the monk's claim that 'There is no Nagasena' is to ask, 'Who guards morality, who meditates, who realizes the paths and fruits and attains Nibbana? Who kills living beings, who takes what is not given...?' The king soon pursues this further by asking why Nagasena himself has 'gone forth' (pursued the life of a monk) (II, 3) and who takes rebirth (assuming that there is such a thing) (II, 4). At this early point in the discussion, the monk's answers seem to side-step or evade the issue that the king is actually raising. Finally (II, 18 and 19), the monk provides the beginnings of an answer to the king's concern by claiming that it is 'mind-and-matter' (also translated, more literally, as 'name-and-form,' sk. *nama-rupa*) that both performs moral actions and takes rebirth. So far, one is inclined to think that such a response sounded

enough like Aristotle's view of the human being as comprised of matter (the body) and form (the soul) so as to place the king on more familiar terrain. However, the monk immediately adds, in reply to a further query of the king, that the 'name-and-form' that performs an action (in one life) is different than the 'name-and-form' that takes rebirth (in the next). Nagasena's point is that it is not the 'name-and-form' (i.e. the individual person) that connects an action and its consequences but, rather, the other way around: it is actions and their consequences that connect one 'name-and-form' to its rebirth (or even to later 'versions' of itself in a single life).

The frequent recurrence of this theme indicates another central difference between Greek and Buddhist views. From a Greek perspective, the fundamental locus of moral praise and blame and of virtue and vice is the soul, as anyone familiar with Plato's *Phaedo* or *Republic* or Aristotle's ethical writings will immediately recognize. Greek moral philosophy, in general, assumes the soul as a morally individuating principle, the basis of 'character,' and tends to ask what virtues should be cultivated or vices avoided, regarding the two as qualities or attributes of the soul. Particular actions are then treated as 'good' or 'evil' to the degree that they tend to establish or support specific virtues or vices possessed by the soul. From a Buddhist perspective, however, this order is almost exactly reversed. Specific actions have definite consequences, these consequences come together in specific configurations, and it is their specific configuration at any given time that allows us both to identify (as well as relate and distinguish) across time a specific 'name-and-form.' Put concisely, then, from the Greek perspective, specific actions result from the condition of souls, what human beings essentially are, while, for the Buddhist, we are what we are at any given time ('name-and-form') is a result of actions that have been performed.

Clearly there is much more of philosophical import to be considered here, but, within the context of the dialogue, we can at least understand how Nagasena's answer represents a plausible response to Milinda's principal concern. For if who or what we are at any given moment (our 'name-and-form') is a product of our past actions, and if certain of those actions have moral qualities, then it makes complete sense to attribute those qualities to the 'name-and-form' whose present existence is formed by them. Likewise, if one accepts the Buddhist account of rebirth, then it is also plausible to regard the consequences of actions as extending beyond the cessation of one 'name-and-form' and continuing so as to constitute another 'name-and-form' standing within the same 'karmic chain' of actions and consequences. In either case, if we *are*, at any given moment, just the ensemble of our actions and their consequences, then we are certainly fully 'morally responsible' for them, perhaps even more so than on the more essentialistic Greek view of the soul and its attributes. To this, Milinda could very earnestly claim, "Well said, revered Nagasena."

Viewed in the overall context of the early part of the dialogue, then, while this may not of itself be sufficient to overcome the king's and monk's quite extreme differences with respect to initial ontological assumptions, it seems to

have removed the major practical objection standing in the way of the king's willingness to attempt a reorientation of his own thinking about more fundamental issues. Whatever else may be the case, the rejection of 'moral nihilism' proves to be common ground for the king and the monk and a basis on which further exchanges can build.

Wisdom

So far, the dialogue has presented an initial confrontation between two opposed ontological beginning points and has, nonetheless, managed to find agreement on the crucial practical issue of moral responsibility. However, oddly intertwined with these discussions is a third theme with distinctively Greek resonances: the nature of wisdom and its relation to other human faculties. Early on (II, 5 and 6), following a brief discussion of rebirth, the king asks, regarding what mental states condition rebirth, whether 'proper attention' (*yoniso manasikara*) is the same as 'wisdom' (*panna*)? The monk responds that they are 'certainly not' the same, since animals have proper attention but lack wisdom. When asked further about their difference, the monk responds that proper attention always has 'comprehension' (or 'examination') as its 'mark' or aim while wisdom has 'cutting off' as its own. The simile offered by Nagasena must have perplexed the king even further, since the monk compares proper attention to a barley reaper's gathering up of the barley into a bunch in one hand and wisdom to his cutting it off with a sickle in the other.

Since several of the king's later questions continue to return to the issue of wisdom, we must assume that this initial exchange on the topic thoroughly perplexed him. And well it might, given a Greek frame of reference. Wisdom (*sophia* in Greek) originally meant 'cleverness or skill in handicraft and art' (as in Homer and Hesiod); later, 'skill in matters of common life, sound judgment, intelligence, practical wisdom'; and, by the time of the great Athenian philosophers, 'learning, wisdom of natural philosophy and mathematics.' (Cf. Liddell and Scott, *A Greek-English Lexicon*.) *Sophia*, as it occurs in the term '*philosophia*,' seems to have retained all these elements, both practical and theoretical, making it the ultimate aim or aspiration of human endeavor. To be a philosopher, as the king seemed to fancy himself, involved both striving to know that which is true and practically applying this knowledge in one's own life. If anything, 'cutting off' of knowledge from life or the mind from the world would be the furthest thing from Greek *sophia*. One must imagine, then, that the monk's initial response was almost incomprehensible to the king, however much the simile may have appealed to him as an image.

Later (II, 12), in the course of some further clarifications of other mental states drawn from *abhidhamma* teachings, the king poses the identical question he had previously asked: 'What is the distinguishing mark of wisdom?' In response, the monk recalls that he had said wisdom (*panna*) was 'cutting off,' but he now emphasizes another characteristic, 'enlightenment' (or

‘illuminating’) (*obhasana*). To illustrate, the monk employs the simile of a lamp dispelling the darkness (ignorance) and causing ‘the effulgence of clear knowledge (*nana*)’ to arise.

Clearly still puzzled about this response, the king, for a third time (II, 15), asks about wisdom, this time referring back to the previous response: ‘Revered Nagasena, has wisdom (*panna*) arisen in him in whom knowledge (*nana*) has arisen?’ This is a telling question, since, within the king’s Greek framework, knowledge (*episteme*) would have been something distinct from wisdom (*sophia*). For both Plato and Aristotle, knowledge (*episteme*) tends always to be ‘about’ something, a ‘knowing that...,’ usually of a particular fact, idea, or skill. Wisdom (*sophia*), however, is more a state a being. While appropriate types of knowledge do form part of Greek wisdom, wisdom generally appears as the superordinate concept, involving not just particular instances or types of knowledge but a sense of their forming a coherent whole together with good judgment in relating or applying them to one’s way of living.

Nagasena’s response categorically rejects such a view: he answers flatly that ‘knowledge (*nana*) is the same as wisdom (*panna*).’ In one of the few passages where the king seems directly to challenge the monk’s response, Milinda presses, ‘But, revered sir, would he in whom knowledge, which is the same as wisdom, has arisen be bewildered or would he not be bewildered?’ The king, that is, is suggesting that if knowledge and wisdom were identical, and if wisdom is the sort of immediate ‘illumination’ that the monk had earlier claimed, then it would be impossible not to know or to be wrong about anything. That is, it would obliterate the Greek distinction between knowledge (*episteme*) and opinion (*doxa*) so fundamental to all Greek philosophy.

In response, Nagasena makes what appears to be an important concession to the king. His answer indicates that, in fact, one in whom wisdom (*panna*) (= knowledge, *nana*) has arisen could still be ‘bewildered’ with respect to some things but not to others. That is, Buddhist wisdom is not omniscience or some universal knowledge of the principles governing all things (as Plato and Aristotle sometimes seem to suggest); the monk concedes that someone possessed of wisdom might still not know many specific things such as those s/he hadn’t learned, places s/he hadn’t been, and words or names s/he had not previously heard. But, with respect to those things that are most important – Nagasena specifically mentions impermanence, suffering, and non-self – one in whom wisdom has arisen will immediately know them to be true. In this sense, then, ‘*panna*’ and ‘*nana*’ are identical: to be wise is to directly know these truths and to directly know these truths is to be wise.

But this is not all: the king has yet one more surprise in store, for the monk contends that once the knowledge appropriate to wisdom is gained, then wisdom has ‘done its task,’ ceases, and only the knowledge of these truths remains. Buddhist wisdom, that is, is not some virtue-like state that subsists in the soul but, rather, a sort of direct insight or illumination that establishes the truth of the *dhamma* for an individual ‘name-and-form’ and then ceases once the knowledge of the *dhamma* has been secured.

The king seems understandably perplexed by this since he asks the monk for not one, as usual, but several different similes to help clarify this teaching. Indeed, this is the longest single exchange in the entire original part of the dialogue. I would suggest that the underlying problem with which Milinda continued to struggle was his reluctance (or perhaps inability) to fully embrace the Buddhist teaching of 'non-self' (*anatmen*). For from a Greek perspective, wisdom was always regarded as a virtue (see, for example, Plato's *Republic*) and virtue was in turn regarded as a quality or property of a soul. While Nagasena tried valiantly to recast wisdom as a sort of impermanent process, an immediate or direct flash of insight, which left behind its 'karmic result' of the knowledge of the *dhamma*, the king was still left to wonder about the 'knower' of this knowledge or the 'thinker' of the *dhammic* thoughts. Unfortunately, the dialogue does not proceed further along this trajectory.

Milinda's 'Conversion'?

The 'Questions of King Milinda' proper (i.e. the first three books) ends on a mutually congratulatory note. Both the king and the monk affirm that 'Everything was properly asked... everything was properly answered.' (III, 26) It is also intimated that the king at least considered 'going forth' as a result of this discussion, but concluded that 'I would not live long, so many are my enemies.' However, at the beginning of Book IV (probably a later addition), we meet Milinda again, this time with a shaved head, donning saffron robes, and undertaking a week-long retreat. At the end of his retreat, he asks to see Nagasena again, but this time in private, without the retinues of either, to continue their discussions, which turn out to be 'higher *dhamma* teachings' involving apparent 'dilemmas' arising within Buddhist doctrine. The style also changes markedly, abandoning the stiff formality of the earlier discussion and the frequent interventions of the *abhidhamma* framework, and adopting a more free-flowing exchange of questions and responses between interlocutors for whom the most basic questions treated in the first major section seem already settled.

What, then, are we to make of the king's apparent 'conversion,' given the rather clear reservations that he voiced in the earlier exchanges? Apart from the king's polite though formulaic replies of 'You are dexterous' or 'Well spoken' that conclude the various earlier sections, there is little textual evidence that the king really understood or, if he did, accepted the monk's responses. And yet, there is some strong historical evidence that the historical King Milinda did, in fact, accept or at least place his personal authority behind the Buddhist way of life. Unfortunately, missing between the end of Book III and the beginning of Book IV is an account of how the king's original exchange with the monk led him, upon reflection, to decide to 'go forth,' and this is not retrospectively supplied in the later additions to the text.

We can therefore only speculate, on the basis of what we have before us in the original text, as to what may have influenced Milinda to abandon his Greek intellectual reservations and set out on the Buddhist path. We should first recall that Milinda, intellectually inquisitive as he was, was not himself a philosopher but a warrior king. He was, therefore, likely not as attached to the more ‘theoretical’ aspects of the Greek philosophical tradition as someone trained in philosophy might be. As we saw, perhaps his principal initial reservation was the practical point of whether the Buddhist doctrine of ‘non-self’ was compatible with moral responsibility. On this score, the monk’s responses to his early questions about this not only seemed to satisfy him (since they do not appear again in the discussion, as other points do), but perhaps even served to convince him that the ‘karmic constitution of name-and-form’ implied an even stricter standard of moral responsibility than did the Greek view. The king was then able to follow the monk’s teachings with a more open mind, since it removed the main obstacle for the king to viewing the idea of ‘non-self’ as at least plausible and worth further exploration. This is not to say that, particularly with respect to the questions of rebirth, knowledge, and wisdom, the king was completely satisfied with the monk’s responses, and he, in fact, occasionally returns to these in later additions to the text. What Milinda comes to understand, perhaps, is that he had approached the entire discussion as an intellectual exercise and in the process discovered that the heart of the Buddhist view concerned an entire way of life, not just a set of intellectual or philosophical theses.

For his part, Nagasena’s responses never required the king to sacrifice his Greek sense of intellectual curiosity and philosophical exploration, while, at the same time, the monk was subtly nudging Milinda toward deciding what was genuinely important in his own questions and offering an at least plausible alternative viewpoint. And this, perhaps, is the key to the appeal of Buddhism, both then and now, in the face of Greek and later Western philosophy: it never requires philosophical curiosity to be suppressed, it focuses upon an entire way of life rather than specific intellectual points, and it presents a sophisticated enough account to at least constitute a plausible alternative.

The point is that, in the end, Milinda did not have to cease being Greek or instantly abandon his most fundamental assumptions to ‘go forth’; rather, perhaps, he came to see that ‘conversion,’ too, was not a ‘state of the soul’ but a process of inquiry supported by practice. And, we must imagine, the Buddhist author of the *Milindapanha* was aware of this as well in depicting the ‘conversion’ of the great Greek king.

The more general conclusion to be drawn is that, while unmistakably Hellenic ideas and themes appear in the dialogue through Milinda’s questions and responses, the encounter is not, in the end, an equally weighted and mutual exchange of competing views. Rather, it is a decidedly Buddhist presentation of an at least practical, if not fully intellectual, victory of the *dhamma* over what any educated Buddhist of the time would have regarded as the ‘false tenets’ of Greek philosophy. Viewed in this way, the *Milindapanha* might be fairly regarded as a dramatic literary portrayal of what is known about actual

*The Milindapanha ('Questions of King Milinda'):
An Early Encounter between Greek and Buddhist Philosophy*

historical circumstances: that resident Greeks discovered enough that was reasonable in Buddhist thought and culture to allow considerable practical accommodation, even as differences over broader philosophical issues and between theoretical frameworks remained.

Bibliographic Note

There are three major English translations of the *Milindapanha*: two of the entire work, by T. W. Rhys Davids (1890) and I. B. Horner (1969), and an abridgement, based on the Horner translation, by N. K. G. Mendis (1993). In this essay, I cite the sections in the latter text, which contains cross-references to the earlier texts. For more detailed background information on the original text, historical context, etc., refer to the extensive introductions of both the Rhys Davids and Horner translations. Of the very few essays dealing with 'Greek ideas' in the *Milindapanha*, especially helpful is J. F. Roccasalvo (1980). 'Greek and Buddhist Wisdom: An Encounter between East and West.' *International Philosophical Quarterly* XX(1), Issue No. 77: 73-85.

CHAPTER SEVEN

Plato on Truth and Falsehood in the Polis

Anne M. Wiles

In the *Republic*¹ (389b-c), Plato endorses for the guardians of his ideal polis the principle of the medicinal lie. The principle can be briefly stated as: ‘It is appropriate for the guardians to lie to the citizens for the benefit of the polis, but it is subversive and destructive for anyone else in the polis to lie.’ Given that the ideal polis Plato is describing is a just state, a morally healthy state, some writers with a post-Kantian moral sensibility have criticized his position as being both inconsistent and dishonest.²

The argument of this paper is that Plato’s position is neither dishonest nor inconsistent, but that when adequately understood in its context, both in the *Republic*, and in the wider context of the other political dialogues, the *Statesman* and the *Laws*, his account of truth and falsehood offers valuable insights into the nature of both and their role in the moral life of the individual and state. In order to concentrate on the positive account of truth and falsehood found primarily in the *Republic* and the *Laws*, this paper will treat only briefly the criticisms of Plato’s position.

To understand medicinal lying in its proper perspective, we must keep in mind certain obvious, but often overlooked, aspects of the *Republic*. First, the *Republic* is an imaginative construct and as Plato (592 a-b) suggests, the ideal does not and perhaps cannot exist anywhere as an historical state. What is important is to understand the principle of medicinal lying so that one can reflect on whether it can be applied in various moral contexts, and if so, the criteria for applying it. The consideration of possible contexts of application is, of course, beyond the scope of this paper.

Second, the justification for the medicinal lie occurs within the wider discussion of the definition of justice. A fundamental insight is that justice must give to each person his due and that persons differ widely in their native

¹All translations are from *The Collected Dialogues of Plato* (1960), eds. Edith Hamilton and Huntington Cairns (Princeton, New Jersey: Princeton University Press). Subsequent references in the paper to passages from the dialogues are given in the body of the text by classical reference number.

²For example, Gregory Vlastos (1977) ‘The Theory of Social Justice within the Polis in Plato’s *Republic*,’ *Interpretations of Plato*, ed. H.F. North (Leiden, Netherlands: E.J.Brill) questions whether a social system which entails the duty of lying deserves to survive, and Karl Popper (1966) *The Open Society and its Enemies*, Vol 1 (Princeton: Princeton University Press), p.138, sees the principle of medicinal lying recommended by Plato as a principle of ‘political expediency’ and ‘totalitarian morality.’

capabilities. Also central to the account of justice is the metaphysics of the person, that a human being is a composite of a reasoning capacity, *thumos* (emotions or passion) and appetite. Justice within a person is a harmony in which each of these aspects is doing its own job well, so that reason is giving good rules for the conduct of life, with the *thumos* supporting the dictates of reason and the appetites obeying the rules. Similarly, in the just state the reasoning function is making good laws, the protective function ensuring they are carried out, and the productive function following the laws. In the *Republic*, the rulers have a wisdom that is lacking in the other citizens. The philosopher-king, qua philosopher, is depicted as a lover of *aletheia* (truth or realities), as knowing both what should be achieved in the polis and how to achieve it, and, qua ruler, as having the power to put in place the means to achieve it.

The distinction in the *Republic* (382 b-d) between two kinds of falsehood is pivotal for understanding Plato's views on truth and falsehood. Essential falsehood, which is ignorance or deception in the soul about realities, is loathed by both gods and by men. Falsehood in words is loathed by the gods but serviceable to men in three instances when it may be used without deserving our abhorrence: against enemies, against *philoï* to avert some evil arising from their madness or folly, and when used in edifying tales.

Thus, there are two types of falsehood, the veritable lie and the verbal lie. There are also two corresponding types of truth: essential or veritable truth when the soul apprehends *aletheia* (truth or reality); and truth in words or verbal truth that occurs when words accurately reflect how one thing is related to another.

Love of *aletheia* is characteristic of the philosopher, but to achieve clarity about truth and reality is a condition not readily attainable. Given the differences among persons in native ability, clearly indicated by Plato throughout the *Republic*, and captured in the Myth of the Metals (414b-415e), it is also not a condition that can be reached by everyone. People differ drastically in wisdom, in their understanding of the nature of things and of how healthy change can best be effected to bring a person or a state closer to the ideal of justice. This is a truth Plato recognizes, and it is a truth still.

Wisdom alone is not enough to bring about a just state. One may be wise and willing to help establish justice and yet fail, as Plato discovered in Syracuse, where the long delayed and inadequate education of Dionysius II and political intrigues defeated Plato's efforts. Nonetheless, the principle is correct: people with native ability, given appropriate education, are wiser than others, and part of their wisdom is knowing the character of persons and situations.

In Book 1 of the *Republic*, there is an adumbration of the nature of the medicinal lie and its relation to the principle of justice. Here (331c), Cephalus and Socrates agree that truth-telling and paying back what one has received do not define justice since each of these actions may sometimes be just and sometimes unjust. Verbal truth, in and of itself, is not a fundamental moral principle; it must be further evaluated in light of the principle of justice. In brief, verbal lying, being context dependent, has an inconstant value and justice as a moral principle trumps truth telling.

In the *Statesman*, as in the *Republic*, the statesman is said to be analogous to the physician. Plato claims that what is true of the physician is also true of the statesman:

So long as they control our health on a scientific basis they may purge and reduce us or they may build us up, but they still remain doctors. The one essential condition is that they act for the good of our bodies to make them better instead of worse, and treat men's ailments in every case as healers acting to preserve life. We must insist that in this disinterested scientific ability we see the distinguishing mark of true authority in medicine and of true authority everywhere else as well (Statesman, 293 b-c).

In the polis, the medicinal lie must be administered on the basis of scientific knowledge, i.e., by one who understands the nature of health in a state and how to obtain it. A corollary is that it should not be administered by one who does not have a proper knowledge of statesmanship, and therefore should be administered by no one except the rulers. Hence the inequality by virtue of which rulers may, under certain conditions, lie to citizens, but not vice versa. The lie must be such as to preserve the state and make it (morally) better, not worse.

Plato describes two cases in which he thinks the use of the medicinal lie would be justifiable. In the first case (459c-460a), concerning marriage regulations, he specifically invokes the principle of the medicinal lie, 'It seems likely that our rulers will have to make considerable use of falsehood and deception for the benefit of their subjects. We said, I believe, that the use of that sort of thing was in the category of medicine.' In order to achieve as many marriages as possible between the best people and the fewest between the worst with the least possible dissension, the rulers are to determine how many marriages there are to be and which persons are to marry, but are to use a system of 'ingenious' lots so that 'the inferior man at each conjugation may blame chance and not the rulers.'

In the second case (414b-415e), Plato recommends that the rulers contrive a 'noble' lie, which, if it does not persuade the rulers, might at least persuade the rest of the city. The story, or myth, to be told is that all the citizens are children of the same mother, Earth. Although they are siblings, God mingled gold in the generation of the rulers, silver in the auxiliaries, and iron and brass in the farmers and other craftsmen as a consequence of which their natural abilities differ. Each person should be assigned status in the state on the basis of his nature. Included in the myth is the oracle that 'the state shall be overthrown when the man of iron or brass is its guardian.' Such a story, if believed in later generations, would have the good effect of making the citizens more inclined to care for the state and for one another.

Critics of Plato's view have centered especially on the ingenious marriage lots, a device which some have claimed violates Plato's own criteria of when a verbal lie is justified: it is not an edifying fable, it is not used against enemies,

and the young guardians with whom it is used are neither mad nor foolish. Now, while it is true that the young guardians are neither mad nor foolish, they are yet young guardians who lack experience and in whom the sense of honor is strong. If some among them were told outright they were less worthy than others to breed with the best mates and that they should produce fewer children, this could understandably lead to their resenting both those making the regulations and those more favored. The result would be disharmony and strife within the guardian class. This would not be for the advantage of the polis, which advantage we recall, is the only reason statesmen are permitted to lie. On the other hand, the marriage lots device will let those not allowed to marry or procreate believe this to be a result of chance, the luck of the draw, and not a deliberate decision that highlights their (relative) inferiority. The device, a medicinal lie, prevents unnecessary resentment and disharmony among the guardians, and being for the advantage of the state, is therefore, on Plato's own principles, justifiable.¹

In the context of the *Republic*, the medicinal lie would be justifiable if and only if it is instrumental in producing an essential harmony in the state (social justice) and personal justice in the individual. Education develops this harmony. Edifying fables, songs, and other types of 'enchantments,' for example, the preambles in the *Laws*, are all components of the musical paideia which persuades citizens to learn to like what the law enjoins and dislike what it forbids.

The noble lie clearly falls under the category of edifying fables. Music is education for the soul and is to begin, in the form of tales or fables, earlier than gymnastic. The educator will make use of one sort of false tale, namely, the fable, which 'taken as a whole is false, but which contains truth also' (*Republic*, 377 a) and will scrupulously avoid another sort of false tale, namely the sort which Homer, Hesiod and other poets relate to us containing false pictures of the gods. The lie is 'not a pretty one when anyone images badly in his speech the true nature of gods and heroes, like a painter whose portraits bear no resemblance to his models' (377 d-e). Fables of the first sort are familiar to anyone who has read Aesop. Fables of the second sort should be banned because they are false in every way, and can do moral harm to one who hears and accepts them.

There is more to edifying fables and other methods of persuasion than first meets the eye. Their use in education of the moral sentiments rests ultimately on the metaphysics of the person. A person is not a disembodied intellect but a delicately balanced composite of both reasoning and affective aspects. As Plato shows in the third speech in the *Phaedrus* (243 e-257 b), the most effective persuasion is that which appeals to the reason by stating what is true and to the emotions by presenting the truth beautifully.

¹For an elaboration of this rationale, and for a review of criticisms of Plato's views on lying, see Thomas C. Brickhouse and Nicholas D. Smith (1983) 'Justice and Dishonesty in Plato's *Republic*,' *Southern Journal of Philosophy*, Vol.21, pp.79-96.

It is on this basis that Plato in the *Laws*¹ speaks frequently of ‘persuading,’ ‘charming, or ‘enchanted’² the citizens in order to produce a concord between reason and emotion. Plato (781 b) comments on persuasion, ‘...the laws’ method will be partly persuasion and partly (when they have to deal with characters that defy persuasion) compulsion and chastisement...’ Karl Popper sees this persuasion as ‘largely lying propaganda,’³ but this interpretation is flawed because the ‘charms’ which Plato is recommending in this passage and in the passages immediately following are not lies; they are merely attractive presentations of the truth.

Concerning persuasion, the Athenian (*Laws* 664 a) declares that the legislator can persuade the young of anything, ‘...the youthful mind will be persuaded of anything, if one will take the trouble to persuade it,’ Even if, *per impossible*,⁴ there could be a strong argument showing that for the individual who gets away with injustice, the unjust life is more pleasant than the just one, the legislator could still consistently recommend that the just life be *presented as* the happy and pleasant life. But why? Because believing that the just life is a pleasant and happy life will induce the citizens to practice justice thus, on the whole, bringing about the greatest good and happiness to the state and to the citizen.

In introducing the three choirs which are to ‘enchant the souls of children,’ (*Laws*, 664 b-c), the Athenian says,

I maintain that our choruses—all three of them—should charm the souls of the children while still young and tender, and uphold all the admirable doctrines we have already formulated, and any we may formulate in the future. We must insist, as the central point of these doctrines, that the gods say the best life does in fact bring most pleasure. If we do that we shall be telling the plain truth, and we shall convince those whom we have to convince more effectively than if we advanced any other doctrine.

¹For example at *Laws*, 659 d-e; See also 664 b, 665 c, 670 c, 671 a, 773 d, 812 c, 873 e, 887 d, 903 b and 944 b.

²Glenn Morrow (1953) ‘Plato’s Conception of Persuasion,’ *Philosophical Review* (26), pp.238-239, discusses the term ‘enchantment’ and its connection with education: ‘Education [in the *Laws*] is essentially a training of the sentiments, from earliest years onward, into accord with the standards set by the law (653 b-c). But in the later work we are given a clearer view of what Plato thought this training involves. He calls it explicitly and repeatedly a process of “enchantment”.... ‘In order to produce this effect, chants [*odai*] appear to have been invented, which are really enchantments [*epodai*], and are designed to implant the harmony of which we speak (659d-e). But the word, *epodai*, and its cognates appear with notable frequency in all later discussions of education, and we can only conclude that Plato is deliberately emphasizing a definite technique. And this is strange, for *epodai* are most commonly connected, elsewhere in Plato and in Greek writers generally, with magic or sorcery. They are the spells with which the sorcerer charms snakes, or drives away diseases, or averts divine wrath.’

³Popper, *The Open Society and Its Enemies*, pp.270-271, notes 5-10, especially note 5.

⁴This sort of case is considered at *Laws*, 663 e.

The remaining comments we find in the *Laws* on lying condemn it. In the preamble to commercial law (916e-917), the citizens are enjoined not to cheat each other, not to swear false oaths, and not to lie to superiors. They should not break their word or contract (920 e- 921 a). There are also injunctions against bearing false witness in a court proceeding (937c). Finally, Plato quotes Hesiod, this time with approval, ‘Justice is the daughter of Respect and both are the natural scourges of falsehood’ (916 e).

In the persuasive, but not lying, general preamble to the legal code, Plato moreover provides a beautiful tribute to truth:

Truth heads the list of all things good, for gods and men alike. Let anyone who intends to be happy and blessed be its partner from the start, so that he may live as much of his life as possible a man of truth. You can trust a man like that, but not the man who is fond of telling deliberate lies (and anyone who is happy to go on producing falsehoods in ignorance of the truth is an idiot). Neither state is anything to envy

This encomium does not say that one must always tell the truth; it says that anyone who wants to be happy and blessed should live as much of his life as possible as a person of truth. The point is that, on the whole and in general, telling the truth is morally and socially preferable to telling a lie. If we love essential truth, we will abhor and try to avoid the essential falsehood, or ignorance of soul. This love of *aletheia* will help bring about one’s own moral and intellectual development. But telling the (verbal) truth also has a positive social value. One trusts a person who tells the truth, and trust is an essential element in uniting persons to form a stable social unit.

When all these threads are pulled together, we are left with a picture of statesmen who love *aletheia*, and who, while they must make considerable use of the verbal (medicinal) lie, are yet not habitual liars, i.e., they do not lie because they are fond of lying, but in every instance only for the good of the state. They recognize and encourage truth-telling as essential among citizens carrying out various legal, social and contractual obligations. They condemn any lie that presents the gods or heroes as evil.

In contemporary moral discourse there has been a tendency to treat all instances of lying as morally objectionable. The aversion most of us feel to lying has been justified on the basis that there is something inherently wrong in lying. Kant, for example, holds that it is always a duty to oneself to tell the truth since, ‘A lie is an abandonment, or as it were, an annihilation of the dignity of man.’¹ Kant’s well-known position that telling a lie is inherently and always morally wrong, a violation of the categorical imperative, takes truth telling to be a moral absolute, whereas Plato, as is evident in these comments from the *Laws*, holds that in most cases and on the whole, lying is to be

¹Immanuel Kant (1964)*The Doctrine of Virtue*. trans. Mary J.Gregor (New York: Harper and Row) p.92.

condemned, but not in every case. For Plato, justice is the fundamental or cardinal principle, and under this principle, in certain rare cases the medicinal use of a lie by wise rulers may be morally justifiable.¹

One might wonder whether the modern critics who emphasize the “dishonesty” of Plato’s view on lying are falling into the very fault that Plato identifies and describes in the *Republic*? ‘Just as too great a devotion to either music or gymnastic causes an undesirable imbalance in the soul,’ he says, ‘so too, in respect of truth, we shall regard as maimed in precisely the same way that soul that hates the voluntary lie and is greatly angered by it in others, but cheerfully accepts the involuntary falsehood and is not distressed when convicted of lack of knowledge’ (535 e).

The judicious Sidgwick, in *The Methods of Ethics*² observes that the absolutist position that holds a lie is never justifiable is difficult to maintain for there are circumstances under which honorable persons would think it right to lie. Plato also takes this position. In the political dialogues examined here, he offers a nuanced treatment of the topic by distinguishing different types of truth and emphasizing how one can be persuaded to love truth. While his position cannot, strictly speaking, be classified as either utilitarian or consequentialist, it nonetheless allows for important contextual features to be part of the moral calculation. Clearly one should love essential truth, and verbal truth as a general rule; the difficult task, for those not perfectly wise, is to determine when the verbal truth should not be told.

¹For a more thorough treatment and evaluation of these opposing views, see Anne M. Wiles (1988) ‘Lying: Its Inconstant Value,’ *Southern Journal of Philosophy* (26) pp.275-284.

²Henry Sidgwick (1907) *The Methods of Ethics* (London: Macmillan&Co.) pp.312-319, discusses the so-called duty of veracity. .

Modern

CHAPTER EIGHT

Locke on Personal Identity: The Form of the Self

Siyaves Azeri

Both in the metaphysical and epistemological spheres, Locke fought on two main fronts. On the one hand, he had to deal with the remnants of Scholastic Aristotelianism. On the other hand, along with Newton's *Principia*, he addressed the issues which later came to be known as the battle between rationalists and empiricists. He dealt with Neo-Platonist epistemology that had been elaborated by the mechanistic approaches of Cartesians; the latter viewed knowledge in terms of a match between human and divine ideas (McCann 1999, 63, 84).

We are furnished with the ideas through experience and we form our knowledge via experience too. Experience is that upon which 'all our knowledge is founded; and from that it ultimately derives itself' (Locke 1975 104). Mind is determined from within itself, yet the determination it brings about cannot be called innate, because such a development happens as a reaction to externally based impressions, and this is how the mind becomes aware of itself for the first time. There are two aspects of reality toward which experience is directed: the external, sensible object and the internal operations of the mind. For Locke, there is no qualitative difference between the knowledge of external things that is based upon sense-data and the knowledge or the awareness of mind (the self) that is founded upon the so-called operations of this mind. Commenting on the experimental nature of knowledge according to empiricism Yolton states:

The empiricist program has been designed to show that all conscious experience 'come from' unconscious encounters with environment, and that all intellectual contents (concepts, ideas) derive from some conscious experimental component (1963/1968, 40).

Locke's claim is directly opposed to the Cartesian argument that the sensations that are caused by a piece of melting wax require interpretation by the intellect via the innate, non-sensory idea of matter. Another important difference between Locke and Descartes is the conception of our awareness of the operations of our minds (i.e., self-reflection). According to Descartes we can have explicit access to such innate ideas (such as substance, duration, etc.) through self-reflection. For Locke, by contrast, self-reflection is simply a part of experience.

External experience, i.e., sensations or sense-data, has a logical as well as physical priority over self-reflection, or the mind's awareness of its own operations. Self-awareness follows from, and comes after, the impressions that are imposed upon the human mind by external entities. The concepts of the mind arising from self-reflection are not innate, but are the consequences of the ideas of the external things that are printed upon the mind starting from birth (Locke 1975, 106). Humans become aware of the outside world, i.e., they form ideas based upon sense-data prior to becoming conscious of their inner impressions. They become aware of the external sensations prior to the operations of their minds which include contemplation. Self-awareness, therefore, requires that the mind be acquainted with the world of objects that are the exterior activity and the affections of the mind. Mind has to operate externally so that the consciousness of such operations arises later.¹

Locke's criticism of the Cartesian identification between soul and thinking anticipates the Kantian notion of the transcendental self (i.e., that awareness which always accompanies the "I think"). He is aware that the soul cannot think perpetually, just as it cannot perceive so. The relation between the soul and thinking and between the soul and perceiving is not similar to the relation between the body and extension, where extension is the essence of spatiality. Rather, it is like the relation between the body and motion. (Apparently, Locke identifies motion with mechanical displacement only.) 'The perception of the *Ideas* being to the soul, what motion is to the body, not its Essence, but one of its Operations' (Locke 1975, 107). The soul is not always active just as body is not always in a perpetual state of motion. To be (a person) is not to think but to be conscious of one's thoughts (as well as one's perceptions and other operations of the mind). Contrary to Descartes, Locke distinguishes between existence and the state of being conscious of all that exist.

The Lockean consideration of the process of thinking suggests that the mind is not a ready-made substance that has certain properties and/or qualifications. Rather, it appears to be an entity that evolves towards a process of understanding. It is by permanent proximity of mind to sense-data that the mind starts to distinguish the familiar from the non-familiar. Through continuous contact with sense-data, a person forms the idea of objects and of other people by receiving their impressions upon his mind. At the level of experiencing external objects and simple ideas then, the mind is passive. It cannot avoid the impressions that affect it just as a mirror cannot avoid reflecting the object opposing it (Locke 1975, 118).² It is through awareness of this external source of knowledge that humans form a second source, namely

¹The soul does not think before senses have furnished it with ideas to think on. The mind first and foremost is involved with situations caused by external objects. The mind employs itself in these operations which can be called perceptions, remembering, consideration, reasoning, etc. (Locke 1975 117).

²H. H. Pierce, writes: '[It is] historically false that the empiricists thought the human mind passive. It would be more just to criticize them for making it more active than it can possibly be' (*Thinking and Experience*, p. 199, note1, quoted at Yolton 1963/1968, 41).

self-reflection as awareness of the operations of the mind. In this latter case, the mind is not only affected but is active.¹

Memory is of central importance to form knowledge. ‘Memory is the storehouse or our ideas’ (Locke 1975, 150). Memory is the constant awareness of ideas by virtue of repeating and fixing them in the mind. Yet memories are in a constant process of dissolution and weakening, and unless the mind has ideas repeatedly impressed upon it, it will lose them (Locke 1975, 151). Since Locke ends up relating personal identity to consciousness and to memory, it is possible to say that, if understanding is not repeatedly and constantly presented with the idea of self-identity, it would lose this idea.

To understand what personal identity consists in, we must understand what the word ‘person’ designates. A person is a ‘thinking intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing in different times and places’ (Locke 1975, 335). This he does only with consciousness, which is inseparable from thinking; indeed, it is essential to it. We are our own selves through the knowledge, that is, the awareness or consciousness of our acts of perception and sensation. The question here is not whether the self resides in the same or in different substances, because, the core of personal identity is consciousness. ‘A person is a single center of consciousness’ (Atherton 1983, 274). The limit of such an identity is as far as the memory of a person can reach back in time. As far as it can reach to those thoughts and feelings of the past, so becomes these same thoughts and feelings part of personal identity; and so in this way the person as he was in the past is one with the person as he is in the present.

However, the question may be raised that no one ever has a complete memory of his past; therefore one’s personal identity is subject to being interrupted due to such lack of completeness. This question, however, tends to equate personal identity with the substantial thinking thing, hence falls short of understanding the nature of personal identity. As was mentioned earlier, according to Locke, personal identity does not consist in substance, be it matter or mind, but in consciousness (and memory) (Locke 1975, 336). Although Locke uses the terms “memory” and “consciousness” interchangeably, we

¹Locke has shown how all ideas arise *after* experience, that is after the organism has encountered the environment and been stimulated into neurophysiological and mental activity.’ He never ‘claimed that ideas arise in the organism in the absence of mental operations, (although some ideas) require relatively few and simple mental operations’ (Yolton 1963/1968, 48-9). However, Yolton’s consideration of activity of mind is problematic and distorts Locke’s approach to passivity and activity of the mind. Passivity of the mind, according to Locke, does not correspond to lack of neurophysiological activities, just as the mind’s activity does not simply mean the presence of such processes. The mind’s passivity, with regard to initial experiences that involve external objects, signifies lack of consciousness and not lack of mental activity. Yolton fails to distinguish between different senses in which Locke has used the word “idea.” Locke’s notion of idea stands not only for sensory items but also for intellectual items namely thoughts or concepts. This is not an ambiguity; rather Locke “holds as a matter of theory that the mental items that come into mind, raw, in sense perception are – after a certain kind of processing—the very items that constitute the basic materials of thinking, believing, and the like” (Bennett 1994, 91).

should be aware that there is a difference between the two.¹ It is not thinking which makes me the self that I am; thought is not the essence of self. It was mentioned earlier that, for Locke, thinking is an activity of the mind. Memory, too, is another form of the mind's activity; it is the recurrence of ideas without perception, since it is a mode of thinking. Consciousness, on the other hand, is that which accompanies thinking. It represents the state of awareness with regard to the operations of the mind. Locke states: 'It being impossible for anyone to perceive, without perceiving that he does perceive' (1975, 335). Hence, consciousness should accompany the act of remembering, too.² As Hegel states, 'Locke maintains thought to be existent in consciousness as conscious thought, and thus brings it forward as a fact in his experience, that we do not always think' (Hegel 1995, 305).

Reducing consciousness to mere memory-ideas results in forcing Locke to go back to a position he is critical of, namely, identification of mind with its contemplative activity, be it thought or memory. The accompanying consciousness can be formulated as the general form of the mind's activities in contrast to the content of the mind's operations. As Yolton says, '[t]o be aware is to be aware of some content of the mind' (Yolton 1963/1968, 43). In the case of memory, the memory-ideas form the content and the consciousness of the existence of such ideas; it is what makes possible the apprehension of such ideas as memory-ideas.

Consciousness always accompanies thinking, and 'tis that, that makes everyone to be, what he calls self, and thereby distinguishes himself from all other thinking things (Locke 1975, 335).

Therefore, it is the constancy and identity of the form of memory, i.e., the continuity of consciousness that provides the constancy of self-identity. 'Locke's idea of what preserves personal identity can be understood simply in terms of consciousness by interpreting this as playing a role analogous to life for an organism' (Atherton 1983, 283). Hence, the change in the content of memory cannot result in the disappearance of personal identity. That a "philosophizing" cat is a cat, be it an intelligent one, and not a person is further evidence showing that Locke differentiates between the content and form of consciousness. The idea of man, to which the term "man" is attributed, 'is nothing else but an animal of such a certain *Form*' (Locke 1975, 333, italics

¹Antony Flew suggests a similar point, yet he draws a different consequence from it; he maintains that 'consciousness is not used by Locke clearly and constantly' (1951/1968, 159). On the contrary, he claims that mainly consciousness is equal to memory and remembrance. However, such a reading is reductionist and cannot give a full account of Locke's response to the question of personal identity. There are numerous instances that Locke uses the term 'consciousness' in a completely different sense than what the term 'memory' signifies. For instance, '[t]hinking consists in being conscious that one thinks' where 'Consciousness is the perception of what passes in a Man's own Mind' (Locke 1975, 115).

²McCann's formulation of Lock's theory of personal identity suggests a reading that runs parallel to mine. In the case of plants and animals the notion of life serves as the principle that organizes the parts as the causal basis of the living thing. In the case of personal identity, consciousness serves as such a principle: 'Simply put, consciousness is the life of the persons. Less simply put, consciousness makes for personal identity in just the way life makes for animal or vegetable identity' (McCann 1999, 75-6).

added). Thus, Reid's criticism of Locke, which tries to show that the unavoidable fluctuation of memory would result in the dissolution of identity of the self, heads in the wrong direction, since it has not taken the difference between form and content of memory into account (Reid 1785/1975, 109-110). Reid's syllogism misses the point by reducing consciousness to mere memory-ideas. It is true that what the person remembers at one moment might be different from what that person will remember at another. However, what is constant and what keeps the identity of a person continuous, is not *what* he remembers but the fact that he is *conscious* that *he* remembers. Each and every time a person recalls something the conscious "I," the general name for the self,¹ accompanies this activity of the mind.

Personal identity can subsist in different substances as long as it carries the same consciousness.² The substance might change over time, yet the person would be the same person given the continuous consciousness of his own personality.

Personal identity reaches no further than consciousness reaches. Even if the supposedly immaterial substance or soul of a Socrates or a Plato is claimed to be residing in someone's mind, still personal identity will be limited to those acts and memories presented to the consciousness of this person. 'As far as this consciousness can be extended backwards to any past Action or Thought, so far reaches the identity of the *person*' (Locke 1975, 335). The reason for this is that the acts of consciousness are attached to individual agents and there is always a reflexive aspect to such activities. Since reflexivity is a mode of experience for Locke, such acts of consciousness should always be accompanied by personal acts of perception. Only those activities are part of my personal identity that are part of my personal experience, whether they be reflexive or object oriented perceptions. Those activities, to which I do not have access, even if they are my own activities, do not partake in the formation of personal identity.

Soul or the immaterial substance alone cannot make the same man; yet consciousness unites the experiences and the existences remote in time and space into the same person. Therefore 'whatever has the consciousness of present and past actions, is the same person to whom they both belong' (Locke 1975, 340). Self depends on consciousness. Since it is the unity of past and present existence and experiences, it is subject to pain and pleasure, happiness and misery.

¹Locke clearly states that '[p]erson is the name for the self' (1975, 346).

²'The identity of substance is not required for identity in every case' (Noonan 1978, 345). Therefore, Butler's criticism against Locke misses the point since he suggests substantial sameness as the principle that guarantees identity of the person and supposes that Locke's account of personal identity works upon the same supposition. Butler infers that consciousness of being the same *person* is consciousness of being the same *substance*, or the same property of a substance, where, in the latter case, the constancy and sameness of the property is the sign of constancy and sameness of the substance. The substance of the self, for Butler, is something like the truth, pure and permanent, and perhaps it is the soul. For details, see Butler's "Of Personal Identity" reprinted in *Personal Identity*, ed. John Perry, 99-105.

Self is that conscious thinking thing, which is sensible, or conscious of Pleasure and Pain, Capable of Happiness or Misery, and so is concerned for it self, as far as that consciousness extends (Locke 1975, 341).

Consciousness joins the body and the soul and forms the person. The self-consciousness of that person will result in his selfhood. The limit of the self is the limit of consciousness. Therefore, whatever is in the reach of consciousness, be it in the past or present, is a part of the person.

Loss of memory results in loss of identity; yet one should be aware that the absence of memory does not correspond to the absence of the content of memory-ideas. As it was indicated above, we should distinguish between 'memory' when it stands for memory-ideas or the content of the memory and 'memory' that signifies the formal element or the consciousness. Content of memory that consists of memory-ideas is in fact subject to constant fluctuation. However, regardless of this, we still speak of the identity of a person as being continuous, because memory, in this second sense, as the factor that is responsible for the identity of the person corresponds to a formal element, namely consciousness. Therefore, loss of memory, which is to be held responsible for loss of the identity of the self, indicates loss of consciousness; it is a symptom of being in the state of unconsciousness. Locke's example regarding juridical practices that may exempt the madman from being punished, and furthermore, the English expression, "one is not himself" are in agreement with such a reading. Earlier in his *Essay* Locke speaks of the madman as being someone who makes incorrect inferences and incorrect propositions. He suggests that the madman maintains his identity although the content of his mind is confused whereas, with respect to the idiot, it is non-existent. What the English expression suggests is not a literal lack of memory but a lack of consciousness, or a lack of access to the content of the mind, which is a sign of a lack of communication between the content of the consciousness, that is, the memory-ideas, the acts and habits of the person, and the person himself. Whether consciousness is related to some immaterial substance or not, it determines personal identity. Hence, the diversity and identity of selves does not result from the diversity of substances, but it is determined only by identity of consciousness. Consequently, any substance that is united within the vital unity of our consciousness is a part of our selves. Upon the separation of man and his consciousness, that is to say, the moment that he is not communicating with his consciousness, the latter is no longer a part of the self. It follows that "person" signifies the very conscious self that exists now.

Where-ever a man finds, what he calls himself, there I think another may say is the same person. It is a Forensic Term appropriating Actions and their Merit; and so belongs only to intelligent Agents capable of a Law, and Happiness and Misery (Locke 1975, 346).

Locke's formulation of experience is limited and restricted. He shares the opinion that perceiving is, in one way or another, having a duplicate image of something that enters into the mind. Locke states:

Our senses, conversant about particular sensible Objects, do convey into the Mind, several distinct Perceptions of things.... [T]hey from external objects convey to the mind what produces there those perceptions (Locke 1975, 105).

Although the mind is active when forming complex ideas, yet, in the most general way and on a larger scale, it is passive. The mind simply obeys the orderliness and regulations that are dictated to it by rules of objectivity, as applicable to physical entities. According to Hegel, Locke takes both the particular and the individual as his principle (1995, 296). This being the case, Locke disallows any role for imagination, rejecting it as frivolous and deranged (Locke 1975, 113). Locke takes existence as individual – be it the object or the self—as granted. His method tends to show how substantiality results from subjective, individual perceptions of objects. Lockean experience does not apply concepts to or reflect them onto objectivity; on the contrary, concepts are simply to be derived from a multitude of singular experiences. For Locke, to know something by experience, ideally, means to know, to perceive, to conceive, or to get acquainted with every single instance of that thing. Locke does not regard the active, willful aspect of experience. He simply relies on the empirical.¹ The question is: how do we come to conceptualize individual, distinct experiences as being experience as such?

Although Locke distinguishes experience from abstraction, yet both rely on the senses, i.e., on the impressions that are imprinted on the mind from outside. Simple ideas come directly and immediately from sense-impressions; complex ones, on the other hand, come from the operations of the mind. Complex ideas express something in the thing that is experienced yet they cannot completely exhaust it. Locke proposes that we do not know the essence of things that is supposed to be reflected in simple ideas; what we do know is the form of the things that are inferred from the nominal essence of things, which is mediated through complex ideas. Locke's major concern, as Hegel expresses, is 'to describe the *manner* in which thought accepts what is given to it' (1995, 310); Locke deals with the forms of attaining knowledge, i.e., his major concern is to attain the forms of the things.

¹I am thankful to Pinar Sumer who brought this general shortcoming of the empiricist conception of experience –which is widely borne out by common sense too—into my attention. Cowley, similarly, speaks of this defective feature of empiricism: 'Empiricists have discussed at length how sense-experience should be spoken of and described, and have distinguished the meanings of terms in which we do so, but remarkably few have discussed our sense-experience of language, and this is intimately connected with the fact that, from Hume to Ayer, the sense-experience they discuss is over-whelmingly visual.... Our sense-experience, on this view, does not include our hearing funny stories, threats, songs, voices like saws, loving murmurs, or witty remarks' (1968, 150).

Locke's formalism resonates in his consideration of the notion of personal identity. What makes the identity is the form of the thing in question. What makes a bridge is the form of the bridge that perhaps may be named 'bridgeness,' so is the case with the identity of plants and animals (as the organization of the parts towards nourishing and conserving the whole). The formation of an oak, from the time it is a seed to the point it becomes a tree, indicates that there exists a form of this particular oak tree regardless of its age, which is reflected in the perception of the organization of it.¹ In the case of personal identity, the form is represented by consciousness that accompanies thought and other operations of the mind. Locke is successful in rejecting the existence of innate ideas; however, his own methodology pushes him toward adopting a Platonic-Aristotelian position with regard to the question of identity.² In contrast to Kant who refers the existence of the forms, including the form of consciousness, to the transcendental self, Locke does not supply any explanation about the origin of these forms. It seems that, unless he admits their independent existence, this question must remain unanswered.

Locke revolutionizes our understanding of the notion of the self. Perhaps, he is the first thinker that openly intends to reject that objects and subjects are qualitatively different. He is at pains to show the dependence of the notion of subjectivity upon representation of external objects that are acquired through sense-impressions. The reality of the self, in this view, is deduced from the fact of its being an object of representation. Moreover, conceptualizing the self as a forensic term is another indication of his general tendency to define selfhood and personal identity as external, objective facts. With Locke the self is posited not as something internal, not as some immediate intimacy but as some totality, which is given through senses, and which is subject to reflexive knowledge, i.e., self-awareness or consciousness.

Yet, in absence of a proper notion of human action and activity along with lack of any reference to the social dimension of the notion of the self, he fails to exploit the potentials that are provided by his non-substantialist inclinations. Thus, he oscillates between two poles of Platonism-Aristotelianism and abstract subjective idealism. Defects of mechanical materialism, as well as abstract approach of subjective idealism to the notion of human activity, which Marx criticizes in his *Theses on Feuerbach*, are detectable in Locke's theory of

¹Although Locke's notions of substance and matter are so manifestly unaristotelian, something like Aristotle's *substantial form* holds a prominent place in his thought, at least with respect to living creatures' (Noonan 1978, 344). However, I believe that such substantial forms are traceable back to material (non-living) structures, as is the case of a bridge whose parts are constantly replaced.

²Antony Flew introduces a number of sources of Locke's mistake in formulating the personal identity. The fifth source of his mistake, says Flew, is 'the assumption that there is some real essence of personal identity, that it is possible to produce a definition and a definition furthermore which can guard us against every threat of future linguistic indecision' (1951/1968, 178). Earlier Flew suggests that the root of such a mistake is Locke's "Platonic-Cartesian conviction that people essentially are incorporeal spirits" and that Locke 'takes for granted that people are souls; which, presumably, conceivably could thus transmigrate' (1951/1978, 169).

personal identity.¹ To the extent that Locke is an objectivist regarding the self, that is, to the extent that he defines the self in terms of external impressions, he fails to explain how the notion of the self is produced. In other words, to the extent that the mind is defined with reference to sense-data and as an objective construct, it is explained as passive and is reduced to simple sensory reactions. Thus, it is an enigma how such a passively formed totality acquires consciousness and is differentiated from other totalities that are formed in reaction to other sense-data. For instance, what element differentiates that particular totality, which is called human from another particular totality, such as a dog?

Although, in case of plants and animals, the notion of nutrition as the principle of organization of life functions as an external but non-formal principle of identity, Locke fails to provide any such principle that constitutes personal identity. Therefore, he is obliged to simply assert the existence of such a form, as something objective but ready-made, which in turn pushes him toward Platonism (objective idealism) and Aristotelianism.

On the other hand, to the extent that he emphasizes the subjective element – consciousness—in the formation of the notion of the self, Locke defines human activity as mere contemplative activity. This is to say, to the extent that the mind is supposed to be active, all its activity is considered pure contemplation. In such a case the aforementioned empty, objective form acquires some content. Locke, then, becomes susceptible to the criticism that has been put forward by Reid and Butler. Moreover, consciousness that is supposed to accompany the act of “I think” becomes indistinguishable from “cogito.” So be the case, his criticism of Cartesianism becomes trivial; replacing *cogito* with *consciousness*, i.e., replacing the “I think” with the “I am conscious” appears as an arbitrary replacement. Consequently, Locke is pushed towards Cartesian dualism and idealism.

Unless a socially determined notion of objective human activity is introduced, the aforementioned tension remains unresolved and Locke inevitably oscillates between Platonism and Cartesianism.

References

- Atherton, M. (1983). ‘Locke’s Theory of Personal Identity.’ *Midwest Studies in Philosophy* 8: 273-293.
- Bennet, J. (1971). *Locke, Berkeley, Hume: Critical Themes*. Oxford: Clarendon Press.
- Bennett, J.: 1994, ‘Locke’s Philosophy of Mind.’ In: V. Chappell (ed.), *The Cambridge Companion to Locke*, 89-114. Cambridge: Cambridge University Press.

¹‘The main defect of all hitherto-existing materialism — that of Feuerbach included — is that the Object [*der Gegenstand*], actuality, sensuousness, are conceived only in the form of the object [*Objekts*], or of contemplation [*Anschauung*], but not as human sensuous activity, practice [*Praxis*], not subjectively. Hence it happened that the active side, in opposition to materialism, was developed by idealism — but only abstractly, since, of course, idealism does not know real, sensuous activity as such’ (Marx 1845/2002, 3).

An Anthology of Philosophical Studies

- Cowley, F. (1968). *A Critique of British Empiricism*. London, Melbourne, & Toronto: Macmillan.
- Flew, A. (1951/1968). 'Locke and the Problem of Personal Identity.' In: C. B. Martin & D. M. Armstrong (eds.), *Locke and Berkeley: A Collection of Critical Essays*, 155-178. Notre Dame & London: University of Notre Dame Press.
- Hegel, G. W. F. (1995). *Lectures on the History of Philosophy*. E. S. Haldane & F. H. Simson (trans.). Lincoln & London: University of Nebraska Press.
- Locke, J. (1975). *An Essay Concerning Human Understanding*, P. H. Nidditch (ed.). Oxford: The Clarendon Press.
- Marx, K. (1845/2002). *Theses on Feuerbach*. In: G. Bagaturia & N. Ter-Akopyan (eds.), MECW, Vol. V, 3-5. Moscow: Progress Publishers.
- McCann, E. (1999). 'Locke on Identity: Matter, Life, and Consciousness.' In: M. Atherton (ed.), *The Empiricists: Critical Essays on Locke, Berkeley, and Hume*, 63-88. Lanham, Boulder, New York, & Oxford: Rowman & Littlefield Publishers.
- Noonan, H. (1978). 'Locke on Personal Identity.' *Philosophy* 53(205): 343-351.
- Reid, T. (1785/1975). 'Of Identity.' In: J. Perry (ed.), *Personal Identity*, 107-112. Berkeley: University of California Press.
- Yolton, J. W. (1963/1968). 'Locke's Concept of Experience.' In: C. B. Martin & D. M. Armstrong (eds.), *Locke and Berkeley: A Collection of Critical Essays*, 40-52. Notre Dame & London: University of Notre Dame Press.

CHAPTER NINE

How Cartesian is Cartesian Linguistics?

Christina Behme

Noam Chomsky coined the term ‘Cartesian Linguistics’ and claimed that his views of language acquisition are firmly rooted in Cartesian tradition (1966, 1967, 1975). This claim has been challenged by Lakoff (1969), Miel (1969), Aarsleff (1971), Percival (1972) and defended by Bracken (1984), Leiber (1988) Lightfoot (1999), Smith (1999), McGilvray (2002), Alanen (2003), and Matthews (2006). Here I am exploring whether or not Descartes held views about language that could be at least compatible with Chomsky’s. If it turns out that Descartes’ views are incompatible with Chomsky’s, then the use of ‘Cartesian’ in ‘Cartesian Linguistics’ would be misleading.

For Chomsky the following three claims about language appear to be essential. First, language is species specific; humans are the only species that has language and non-human communication systems are distinctly different. Second, language is domain specific. While language is embedded in the broader context of general intelligence there are certain features of language (e.g., unbounded creativity, recursivity) that are not shared by any other domain. These two points roughly equal what McGilvray (2002) calls ‘linguistic creativity’. Third, language cannot be learned from the evidence available to the language learner (poverty of the stimulus) and requires therefore innate knowledge. Below I will provide textual evidence supporting my hypothesis that while Descartes believed that language is one of the distinguishing features between humans and non-human animals he was neither committed to the view that language is domain specific nor did he seem to hold that language acquisition depends on innate knowledge in a way that was similar to Chomsky’s.

Species Specificity of Language

Many commentators on Descartes (e.g., Kenny, 1968; Rodis-Lewis, 1978; Wilson, 1978; Bracken, 1982; MacDonald Ross, 1988; Cottingham, 1986; Leiber, 1988; Gaukroger, 1995; Baker & Morris, 1996; Rozemond, 1998; Flage&Bonnen, 1999; Des Chene, 2001; Alanen, 2003; Clarke, 2003; Sorell, 2005; Skirry, 2005; Williams, 2005; Brown, 2006; Gombay, 2007; Schmaltz, 2008; Hatfield, 2008) observe that for Descartes one of the distinguishing characteristics of humans is the use of language. This observation is well supported by textual evidence. In *Discourse on Method* Descartes explains:

‘...there are no men... [who] are incapable of arranging various words together and forming an utterance from them in order to make their thoughts understood; whereas there is no other animal... that can do the like’ (CSM I, p. 140). In a letter to Renieri Descartes explains that automatons replicating animals and those replicating human could be distinguished because the former ‘never answer in word or sign, except by chance to questions put to them’ (CSMK III, p. 99). In a letter to Newcastle Descartes insists that ‘spoken words, or other signs that have reference to particular topics without expressing any passions’ (CSMK III, p.303) set humans apart from animals and machines; ‘the use of words, so defined is something peculiar to human beings’ (Ibid.). And, in a letter to More Descartes writes ‘...it has never been observed that any brute animal has attained the perfection of using real speech, that is to say, of indicating by word or sign something relating to thought alone and not to natural impulse’ (CSMK III, p. 366).

Descartes carefully highlights the difference between superficial similarities in communication behaviour on the one hand and true language use on the other. This is important because it accounts for the fact that some animals (e.g., parrots, magpies) have been taught (some) words and (seemingly) use these words to the same ends as humans do. Yet, according to Descartes there is a fundamental difference between word use in birds and humans: ‘If you teach a magpie to say good-day to its mistress when it sees her approach, this can only be by making the utterance of this word the expression of one of its passions. For instance it will be the expression of the hope of eating if it has always been given a tidbit when it says it’ (CSMK III, p. 303). Descartes points out that the magpie does not say ‘good-day’ to its mistress because it knows the meaning of the words but because it has learned that it will be rewarded with a tidbit. This use of words is not any different from the tricks a dog might perform once it has learned that it will be rewarded for such a performance. Possibly the strongest expression of Descartes’ conviction that language is species specific is his short reply to More’s final objections where he states that ‘Infants are in a different case from animals: I should not judge that infants were endowed with minds unless I saw that they were of the same nature as adults; but animals never develop to a point where any certain sign of thought can be detected in them’ (CSMK III, p. 374). This passage is significant because, unlike in the other cases, Descartes cannot provide any empirical evidence for his claim that infants are endowed with minds. Observation would tell us that infants are indistinguishable from animals because they are not able to use language. Nevertheless, Descartes is convinced that their inability to speak is different from that of animals. Animals will never ‘develop to a point’ where they acquire language while infants will grow up to become language users. They are ‘of the same nature as adults’. Therefore it is membership in a class (humans) not currently displayed behaviour that determines whether or not an organism has or will develop the ability to use language. For Descartes it is clear that only humans belong into the category of language users.

In summary, the textual evidence strongly suggests that Descartes believed that language is species specific. However, the fact that language is species specific could be explained in two different ways. In the first scenario humans have a domain specific language faculty while animals lack this faculty. In this case it would be imaginable that an animal, who had implanted an artificial language faculty, would behave in ways that are indistinguishable from a human being. This is essentially what Alan Turing (1950) believed when he conceived of the Turing test: if a machine displays language behaviour that is indistinguishable from that of a human speaker, then this machine is intelligent. In the second scenario language is one of several indicators of general intelligence (thought and reason in Descartes' terminology). In this case it would not be possible to 'construct' a language faculty that is independent of 'general intelligence'. The differences between humans and non-human animals are not merely differences in their ability or inability to use language but more fundamental differences. To bring it to a point: in the first scenario humans are intelligent because they have language; in the second scenario humans have language because they are intelligent. In the next section I will provide evidence for my claim that Descartes was committed to the second scenario.

Domain Specificity of Language

Chomsky's commitment to a domain specific language faculty is well documented. He stresses that language is learned earlier and with more ease than any other intellectual task (e.g., mathematics), that it appears to be largely independent of the actual input and the overall intelligence and that it seems to be affected by specific types of brain damage that have no effect on 'overall' intelligence. These indicators for a domain specific language faculty are accepted by Chomsky followers (e.g., McGilvray, 2002; Matthews, 2006; Russell, 2004; Smith, 1999; Stainton, 1996). Robert Matthews holds that Descartes and Leibniz attributed 'unlearned knowledge to innate endowment. Learners, they argued, come to the learning task with certain innate, *domain-specific* knowledge that enables them to learn what they do' (Matthews, 2006, p. 82, emphasis added). Even some authors who are critical of Chomsky's interpretation of Descartes hold that the idea of a domain specific language faculty can be traced back to Descartes' writings. For instance, Fiona Cowie suggests that Descartes is committed 'to the view that the mind possesses more inherent structure than is allowed by [empiricists]' and that he believes that this structure is responsible for domain specific (language) learning (Cowie, 1999, p.43).

At first glance Descartes seems to hold such a view; he seems to believe that language can be independent from intelligence (reason). In a letter to Renieri he remarks that automatons replicating animals attempting to imitate our use of language would 'fail more disastrously than the greatest fool' (CSMK III, p. 99), in a letter to Newcastle he refers to 'the speech of madmen

which has reference to particular topics even though it does not follow reason' (CSMK III, p. 303) and in a letter to More he remarks '...all human beings use [language], however stupid and insane they may be...' (CSMK III, p. 366). In *Discourse on Method* the references is equally clear: '...there are no men so dull-witted or stupid - and this even includes madmen - that are incapable of arranging various words together and forming an utterance from them...' (CSM I, p. 140). The verdict appears to be that language can be sustained without a lot of intelligence: 'For it patently requires very little reason to be able to speak' (Ibid.). Seemingly it requires so little intelligence that 'even the stupidest child or ... a child with a defective brain' (Ibid.) can use language. However, Descartes claims at no point, that *no* intelligence (reason) is required for language. In fact when we pay attention to context we see that the purpose of Descartes' examples is to show that the fundamental difference between humans and animals is the possession or lack of a rational mind (soul). So language is not the distinguishing feature but rather a reliable indicator for this feature. This view is confirmed by the text. In *Discourse on Method* Descartes discusses the abilities of animals and observes that (some) animals have the necessary organs to speak and show more skill in some tasks than humans do but nevertheless 'they cannot speak as we do, that is they cannot show that they are thinking...' (CSM I, p. 140). He holds that if the fact that animals excel in some tasks more than humans would indicate that these tasks require intelligence, then these animals should be more intelligent than humans and excel in all tasks. Since this clearly is not the case Descartes concludes that those tasks do not require intelligence:

"...so what [animals] do better does not prove that they have any intelligence, for if it did they would have more intelligence than any of us and would excel us in everything. It proves rather that they have no intelligence at all, and that it is nature which acts in them according to the dispositions of their organs..." (CSM I, p. 141).

Here Descartes stresses that 'nature' allows creatures to excel in some things while intelligence allows us to excel in other things. One of these latter things is using language. Similarly, he stresses in letters to More that 'speech is the only certain *sign* of thought' (CSMK III, p. 366, emphasis added) and again that 'speech, ... alone shows the thought hidden in the body' (CSMK III, p. 374). In the letter to Newcastle Descartes remarks 'that the reason that animals do not speak as we do is not that they lack the organs but that they have no thoughts' (CSM III, p. 303). This would indicate that language does not depend on a special organ (like a language faculty) but that language is the outward expression of thought.

Descartes observed that (at least some) animals do have bodily organs that are sufficiently similar to the bodily organs involved in human speech. Furthermore, he reports on cases in which animals have been taught to produce some words of the human language. This indicates that having the organs dedicated to language production is not sufficient for language use. What is

required is a rational mind and animals lack such a mind. But clearly humans have language because they have rational minds (not the other way around). And minds for Descartes are indivisible: 'we cannot understand a mind except as being indivisible. For we cannot conceive of half a mind' (CSM II, p. 9). This fact is explained in *Meditation Six*:

...minds are utterly indivisible. For when I consider the mind, or myself in so far as I am merely a thinking thing, I am unable to distinguish any parts within myself... As for the faculties of willing, of understanding, of sensory perception and so on, these cannot be termed parts of the mind, since it is one and the same mind that wills, and understands and has sensory perception (CSM II, p. 59).

Descartes does not mention language as a possible faculty here. In fact the two intellectual faculties he mentions (willing and understanding) both use language. While it is possible to claim that a domain specific language faculty could be contained in 'and so on' it appears more plausible to suggest that our linguistic capabilities depend on the mind's essence of rational thought. Descartes stresses repeatedly that the essence of mind (soul) is thought. In *Discourse on Method* he writes that his mind (soul) is 'a substance whose whole essence or nature is simply to think' (CSM I, p. 127). In the conversation with Burman Descartes maintains that 'the mind can never be without thought; it can of course be without this or that thought but it cannot be without *some* thought. In the same way, the body cannot, even for a moment, be without extension' (CSMK III, p. 336, original emphasis). Finally, in a letter to Hyperaspistes Descartes does not endorse an internal structure of the mind but rather the mind's ability to acquire new concepts: 'even though the mind is indivisible, it is none the less capable of acquiring various properties' (CSMK III, p. 196).

In conclusion, while there is some indication that Descartes might endorse a domain specific language faculty the textual evidence for this claim is far from conclusive. It remains possible that Descartes held that humans can acquire and use language because they have rational minds. While language is an essential characteristic of humans (species specific) it is only an indicator of a rational mind not its foundation.

Language Acquisition

For the following discussion it is important to remember that Chomsky (1966) and several of his supporters (e.g., McGilvray 2002; Matthews, 2006) and critiques (e.g., Cowie, 1999) hold that Descartes was committed to a view that language acquisition relies on some form of innate knowledge. They invariably cite Descartes' writings on innate ideas in support of this claim. However, very little attention is paid to what Descartes wrote specifically about the acquisition of language and about the difference between ideas we acquire

in our childhood and later in life. I will cite some of these passages because they could support a different view on language acquisition. One of the most explicit comments about the first acquisition of language occurs in *Search for Truth* where Eudoxus compares the acquisition of scientific knowledge and ‘simple forms of knowledge’: ‘... notice how the sciences differ from those simple forms of knowledge which can be acquired without any process of reasoning, such as *languages*, history, geography and in general *any subject which rests on experience alone*’ (CSM II, p. 403, emphasis added). This is a surprisingly empiricist statement and defenders of the stringent rationalist interpretation of Descartes might dismiss it as inconsistent with his “real” view about language acquisition. However, several other passages might be troubling for the rationalist-nativist interpreter as well.

In a letter to Newcastle Descartes stresses the fundamental difference in language capacities between humans and animals and makes the point that ‘even deaf-mutes *invent special signs* to express their thoughts’ (CSMK, III, p.303, emphasis added). In *Discourse on Method* the same point is made ‘...men born deaf and dumb... *normally invent their own signs* to make themselves understood by those who, being regularly in their company, have the time to learn their language’ (CSM, I, p. 140, emphasis added). If language relies on species-specific innate knowledge, then it seems odd to suggest that at least some people need to invent signs for communication. One could claim that deaf-mutes are an exception not the rule and argue that as members of the human species they have an innate language faculty but their specific handicap prevents the normal maturation/expression of this faculty. However, in a letter to Mersenne Descartes states that ‘words are human *inventions*’ (CSMK III, p. 187, emphasis added) and here he does not refer to handicapped people but he makes a general statement about language. He continues to explain that either one word or several could be used to express ‘the same thing’ (Ibid.) and insists that there is a difference between the words and the ideas they express. Only the latter are innate.

There is further evidence that Descartes believes that language acquisition relies on learning instead of the maturation of an innate language organ. In a letter to Mersenne he asserts that ‘[t]here are only two things *to learn* in any language: the meaning of words and grammar’ (CSMK, III, p. 10, emphasis added) and in a letter to Chanut he explains in some detail how we acquire the meaning of words:

... when we learn a language, we connect the letters or the pronunciation of certain words, which are material things, with their meaning, which are thoughts, so that when we later hear the same words, we conceive the same things, and when we conceive the same things, we remember the same words (CSMK, III, p. 307).

Here Descartes talks about the essential connection between sense perception (objects and sound combinations) and mental concepts for language acquisition. Furthermore, Descartes repeatedly states that language arbitrarily

connects sounds to objects and that we rely on memory when we recall words we have heard previously. In a conversation with Burman he explains:

When for example on hearing that the word “K-I-N-G” signifies supreme power, I commit this to my memory and then subsequently recall the meaning by means of my memory, it must be intellectual memory that makes this possible For there is no relationship between the four letters (K-I-N-G), which would enable me to derive the meaning from the letters. It is intellectual memory that enables me to recall what the letters stand for (CSMK III, pp. 336-7).

This account is quite compatible with any current account of language learning because it requires no faculty more domain specific than ‘intellectual memory’. A similar account is given in *Principles of Philosophy*: ‘...we tie all our concepts to the words used to express them; and when we store the concepts in our memory we always simultaneously store the corresponding words’ (CSM I, p. 220). The fact that we need to store our concepts simultaneously with words could indicate that there is no innate connection between the two. Furthermore, language acquisition is seemingly an empirical affair, because words are learned either by association to objects or by instruction from others.

Finally, Descartes repeatedly points out that the ideas we acquire in our childhood are often ‘confused thoughts’ (CSMK, III, p. 308), based on ‘false preconceptions’ (CSMK, III, p. 233). Usually we learn only much later about the true nature of things and acquire clear and distinct ideas. One example for this difference can be found in the discussion of our different ideas of the sun:

...there are two different ideas of the sun which I find within me. One of them which is acquired as it were from the senses and which is a prime example of an idea which I reckon to come from an external source, makes the sun appear very small. The other idea is based on astronomical reasoning, that is, it is derived from certain notions that are innate in me (or else it is constructed by me in some other way), and this idea shows the sun to be several times larger than the earth. (CSM, II, p. 27)

It is evident that Descartes does not deny that we have the first idea of the sun. Moreover, children probably have *only* this idea. That does not prevent them from using the word “sun” in appropriate circumstances, creatively and stimulus free. Even young children understand that the word “sun” refers to a certain object and they do not need to see the sun every time they talk about it. For everyday communication either idea of the sun will do just fine. A similar observation is expressed in *Search for Truth* when Descartes refers to the learning in early childhood: ‘...he came into the world in ignorance and ... the knowledge he had as a child was based solely on the weak foundation of the senses and the authority of his teachers’ (CSM II, p. 400). Part of the

knowledge acquired in early childhood is knowledge of language. And while Descartes makes it clear that science could not be placed on the shaky foundation of empirical knowledge he does not seem to have similar concerns about language.

In summary, Descartes states that language ‘can be acquired without any process of reasoning... [based] on experience alone’ (CSM, II, p. 403), that we learn language by connecting words with their meanings and remembering later upon encountering words which things they signify and vice versa (CSMK, III, p. 307) at a time when our thoughts are ‘confused’ and based of ‘misconceptions’. While Descartes suggests that later in life we can (and should) acquire a new way of thinking he does not indicate that our language needs to be changed as well.

Conclusion

In this paper I have shown that there are considerable differences between the views of Descartes and Chomsky regarding language acquisition and language use. Based on textual evidence, I defended the claim that Descartes believed that language is species specific. Language is a reliable indicator of thought and thought is only found in creatures that have a rational mind. It is not evident from the text whether or not Descartes would endorse a domain specific language faculty and some of his statements about language acquisition could support a view that is compatible with current theories of data-driven domain general language learning (Redington & Chater, 1997; Sampson, 2002; Reali & Christiansen, 2005). It is clear that I will need to show that my reading of Descartes’ commitments is compatible with his views about ideas in general and innate ideas in particular. An adequate discussion of these views has to be the topic of another paper. Here I only refer to one commentator who claims that for Descartes all ideas (including sensory ideas) “must be formed by an inborn faculty of the mind” (Schmaltz, 1997, p. 36). This faculty is not domain specific but underwrites all intelligent behaviour. Such an interpretation of Descartes’ “innate ideas” would be compatible with the views presented here.

References

- Aarsleff, H. (1971). ‘Cartesian Linguistics’: History or Phantasy? *Language Sciences*, 17, 1-12.
- Alanen, L. (2003). *Descartes’ concept of mind*. Cambridge: Harvard University Press.
- Baker, G., & K. Morris (1996). *Descartes’ Dualism*. London: Routledge.
- Bracken, H. (1982). Chomsky's Variations on a Theme by Descartes. *History of Philosophy*, 181-192.
- Bracken, H. (1984). *Mind and Language: Essays on Descartes and Chomsky*. Dordrecht: Foris Publications.

- Brown, D. (2006). *Descartes and the Passionate Mind*. Cambridge: Cambridge University Press.
- Chomsky, N. (1966). *Cartesian Linguistics. A chapter in the history of Rationalist Thought*. New York: Harper & Row.
- Chomsky, N. (1967). Recent contributions to the theory of innate ideas. *Synthese* 1, 1-11.
- Chomsky, N. (1975). *Cartesian Linguistics*. In: S. Stich (ed.) *Innate ideas*. 89 - 103. Berkeley: University of California Press.
- Clarke, D. (2003). *Descartes's Theory of Mind*. Oxford: Oxford University Press.
- Cottingham, J. (1986). *Descartes*. Oxford: Blackwell.
- Cowie, F. (1999). *What's within? Nativism reconsidered*. Oxford: Oxford University Press.
- Des Chene, D. (2001). *Spirits and Clocks: Machine and Organism in Descartes*. Ithaca: Cornell University Press.
- Doney, W. (1967). Introduction. In: W. Doney (ed.) *Descartes - A collection of critical essays*. 7 - 27. Garden City, NY: Anchor Books.
- Erion, G. (2001). The Cartesian Test for Automatism. *Minds and Machines* 11, 29-39.
- Flage, D. & C. Bonnen (1999). *Descartes and Method*. London: Routledge.
- Gaukroger, S. (1995). *Descartes: An Intellectual Biography*. Oxford: Clarendon Press.
- Gombay, A. (2007). *Descartes*. Oxford: Blackwell Publishing.
- Hatfield, G. (2008). Mental Acts and Mechanistic Psychology in Descartes' Passions. In: N. Robertson, G. McOuat, & T. Vinci, (eds.). *Descartes and the Modern*. 49-71. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Jackson, F. (1982). Epiphenomenal Qualia, *Philosophical Quarterly*, 32, 127-136.
- Jolley, N. (1988). Leibniz and Malebranche on innate ideas. *The Philosophical Review*, 97, 71 - 91.
- Kenny, A. (1967). Descartes on Ideas. In: W. Doney (ed.) *Descartes - A collection of critical essays*. 227 - 249. Garden City, NY: Anchor Books.
- Kenny, A. (1968). *Descartes: A Study of His Philosophy*. New York: Random House.
- Lakoff, R. (1969). Review of H. Brekle (ed.) *Grammaire generale et raisonnee*. *Language*, 45, 343-364.
- Leiber, J. (1988). Cartesian Linguistics? In: The Chomskyan Turn: Generative Linguistics, Mathematics, Philosophy and Psychology International Workshop, Tel Aviv, *Philosophia*, 1988, 309-46.
- MacDonald Ross, G. (1988). Hobbes and Descartes on the relation between language and consciousness. *Synthese*, 75, 217-229.
- Matthew, R. (2006). On the innateness of language. In: R. Stainton, (ed.) *Contemporary debates in cognitive science*. 81-96. Malden, MA: Blackwell.
- McGilvray, J. (2002). Introduction for cybereditions. In: N. Chomsky. *Cartesian Linguistics. A chapter in the history of Rationalist Thought*. 7 - 44. Christchurch, NZ: Cybereditions.
- Miel, J. (1969). Pascal, Port-Royal, and Cartesian Linguistics. *Journal of the History of Ideas*, 30, 261-271.
- Morris, J. (1971). *Descartes Dictionary*. New York: Philosophical Library.
- Percival, K. (1972). On the non-existence of Cartesian linguistics. In: R. Buttler (ed.). *Cartesian Studies*. 137 - 145. Oxford: Basil Blackwell.
- Reali, F., & Christiansen, M. (2005). Uncovering the richness of the stimulus: Structural dependence and indirect statistical evidence. *Cognitive Science*, 29, 1007-1028.
- Redington, M., & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences*, 1, 273-281.

- Rodis-Lewis, G. (1978). Limitations of the mechanical model in the Cartesian conception of the organism. In: M. Hooker (ed.) *Descartes: Critical and Interpretative Essays*. 152 - 170. Baltimore: John Hopkins University Press.
- Rozemond, M. (1998). *Descartes's Dualism*. Cambridge MA: Harvard University Press.
- Russell, J. (2004). *What is language development?* Oxford: Oxford University Press.
- Sampson, G. (2002). Exploring the richness of the stimulus. *The Linguistic Review*, 19, 73-104.
- Schmaltz, T. (1997). Descartes on innate ideas, sensation and scholasticism: the response to Regius. In: M. Steward (ed.) *Studies in Seventeenth-Century Philosophy*. 33-73. Oxford: Clarendon Press.
- Schmaltz, T. (2008). *Descartes on Causation*. Oxford: Oxford University Press.
- Skirry, J. (2005). *Descartes and the Metaphysics of Human Nature*. London: Thoemmes-Continuum Press.
- Smith, N. (1999). *Chomsky: Ideas and ideals*. Cambridge: Cambridge University Press.
- Sorell, T. (2005). *Descartes reinvented*. Cambridge: Cambridge University Press.
- Stainton, R. (1996). *Philosophical perspectives on language*. Peterborough, ON: Broadview.
- Turing, A., 1950, Computing machinery and intelligence, *Mind*, 50, 433-460.
- Williams, B. (2005). *Descartes: the project of pure inquiry*. New York: Routledge.
- Wilson, M. (1978). *Descartes*. London: Routledge and Kegan Paul.
- [CSM I, II; CSMK III] The Philosophical Writings of Descartes, v. I, II, transl. J. Cottingham, R. Stoothoff, & D. Murdoch, and v. III, transl. J. Cottingham, R. Stoothoff, D. Murdoch, & A. Kenny (Cambridge: Cambridge University Press, 1984, 1985, 1991). References are to volume and page number.

CHAPTER TEN

The Conceptualist Analysis of Descartes's Creation Doctrine

Daniel Considine

In several letters to Marin Mersenne, and in his Replies to the Sixth Set of Objections to the *Meditations*,¹ Descartes makes the startling claim that 'eternal truths' – such as metaphysical truths, mathematical truths, and logical truths – are freely created by and depend entirely upon God. Normally, we would think that these kinds of propositions enjoy a necessity and timelessness independently of how God would choose to create them. But Descartes insists that these truths are created *freely*, that God could have done otherwise in creating them. The immediate problem with this contention is: How could something eternal be both created and contingent?

Descartes seems to hold this position because of his insistence that absolutely everything depends on God. To claim that God wills the eternal truths because they are true would be to undermine this omnipotence; it would be to claim that something exists – namely, eternal truths – prior to God's willing, and Descartes is unwilling to stray from his adherence to this omnipotence thesis. The drawback to this position is the untenable consequence that the eternal truths are contingent: if God could have done otherwise in creating eternal truths, then, it seems, they are not necessary. Descartes wants the compatibility of two theses that intuitively appear contradictory: (1) that the eternal truths couldn't have been otherwise than they are, and (2) that God could have chosen otherwise in creating the eternal truths.

How are we to understand the necessity of such truths given that they could have been otherwise? Doesn't this entail an incoherence? For if they are created by God's free will, there is a sense in which they could have been otherwise; and surely, nothing which could have been otherwise is necessary, if necessity is understood as that which cannot have been otherwise. In this essay, I investigate one proposed solution to this tension, the conceptualist analysis of modality as proposed and defended by both Jonathan Bennett (1994) and Timo

¹References to Descartes's works are as follows: 'AT' followed by volume and page number refer to Charles Adam and Paul Tannery (eds.), *Oeuvres de Descartes*, 11 volumes (Paris: Léopold Cerf, 1897-1913); 'CSM' followed by volume and page number refers to John Cottingham, Robert Stoothoff, and Dugald Murdoch (eds.), *The Philosophical Writings of Descartes*, Volumes I and II (Cambridge: Cambridge University Press, 1984 and 1985); 'CSMK' followed by page number refers to Cottingham, Stoothoff, Murdoch, and Anthony Kenny (eds.), *The Philosophical Writings of Descartes*, Volume III (Cambridge: Cambridge University Press, 1991).

Kajamies (1999).¹ This analysis attempts to reconcile the problem between what God could have done and the necessity of the eternal truths by defining modality subjectively. I proceed as follows. First, I outline the conceptualist view, building the best possible case for it. Then I offer several reasons why we should reject it. My main contention is that the view does not comport with pivotal texts which I believe point to an objective interpretation of modality. Finally, I suggest one route to a solution to interpreting the doctrine which incorporates Descartes's theological investigations and the metaphysical foundations of physics.

Conceptualism as an Interpretation

According to the conceptualist analysis, modal terms – namely, necessity, impossibility, and possibility – are strictly features of our minds: they do not describe features of the world.² A proposition or state of affairs as such (whether mathematical, or physical) contains no necessity; only our thought imposes necessity onto it. An early proponent of the conceptualist interpretation, Margaret Wilson, writes, 'Descartes did regard the "necessity" we perceive in mathematical propositions as in some sense and degree a function of the constitution of our minds'.³ And according to Jonathan Bennett, who bases his argument on Wilson's remark, the modal concepts in Descartes's philosophy 'should be understood or analyzed in terms of what does or does not lie within the compass of our ways of thinking'.⁴ Bennett believes that Descartes really intends a subjective meaning for the concepts 'possible', 'necessary', and 'impossible', as such an interpretation dissolves most of the problems of coherence surrounding the doctrine. And this in turn informs the doctrine's principal idea that God establishes the eternal truths of our world, because it interprets that as the claim that God establishes the eternal truths *by* making us unable to conceive of their opposites.

Bennett is concerned to square Descartes's contention that God could have made necessary truths false with the rest of his philosophical theses, such as the truth rule and the argument for the distinction between mind and body. If the view that God could have made necessary truths false entails that our

¹Jonathan Bennett, 'Descartes' Theory of Modality' *Philosophical Review* 103 (1994): 639-667, and Timo Kajamies, 'The Concept of Power and the Eternity of the Eternal Truths in Descartes' *Southern Journal of Philosophy* 37 (1999): 189-200.

²The conceptualist view was first stated by Margaret Wilson in *Descartes* (London: Routledge & Kegan Paul, 1978), pp. 125-131; also by Jacques Bouveresse, 'La théorie du possible chez Descartes' *Revue Internationale de Philosophie* 146 (1983): 293-310; Hidé Ishiguro, 'The Status of Necessity and Impossibility in Descartes' in Amelie Rorty (ed.) *Essays on Descartes's Meditations* (Berkeley: University of California, 1986), 459-471; and Lilli Alanen, 'Descartes, Conceivability, and Logical Modality' in Tamara Horowitz and Gerald J. Massey (eds.) *Thought Experiments in Science and Philosophy* (Savage, Maryland: Rowman & Littlefield, 1991): 65-84, esp. p. 75; but the view gets its fullest development and defense by Jonathan Bennett and, defending Bennett's particular position, Timo Kajamies.

³Wilson (1978), 125.

⁴Bennett (1994), 647.

judgments on the basis of our clear and distinct perceptions of any necessary truth could be false, then Descartes would be undermining his other philosophical arguments simply because of adherence to this thorny doctrine. For example, if the creation doctrine is brought to bear on the cogito argument from the Second Meditation, we could never rely on the truth of the premise¹ that whatever thinks must exist, and so could never get to 'I exist'. But rather than rejecting the creation doctrine as some 'blind spot' which Descartes should have avoided, Bennett seeks to accommodate the doctrine, especially following Descartes's own emphatic endorsement of it to Mersenne in 1630.

Bennett believes his interpretation is the key to settling the problem of incoherence in the texts regarding the creation doctrine. For if modality is understood as a feature of our own minds, we avoid the problem of claims such as 'God could have brought it about that $2+2=5$ ', for these claims would no longer be analyzed in terms of objective possibilities. Bennett points to a passage from Descartes's Second Replies as providing the best evidence for his subjective reading of Cartesian modality:

If by 'possible' you mean what everyone commonly means, namely 'whatever does not conflict with our human concepts', then it is manifest that the nature of God, as I have described it, is possible in this sense, since I supposed it to contain only what, according to our clear and distinct perceptions, must belong to it; and hence it cannot conflict with our concepts. Alternatively, you may well be imagining some other kind of possibility which relates to the object itself; but unless this matches the first sort of possibility it can never be known by the human intellect, and so it does not so much support a denial of God's nature and existence as serve to undermine every other item of human knowledge. (AT VII 150-151; CSM II 107)

This passage indeed appears to be a strong basis for the conceptualist view. Descartes seems to be rejecting a definition of 'possible' which acknowledges possibilities outside the mind. Bennett and Kajamies both interpret this passage as claiming that modality relating to the objects themselves is a 'contrivance'. Whereas CSM translate the Latin verb *fingere* as 'imagining', Bennett uses the stronger 'inventing', suggesting that the proposed definition of possibility has no basis in reality. As Bennett says, for Descartes, 'all modal truths are at bottom truths about what we can conceive'.² Possibilities and necessities that are postulated divorced from our conception of them are simply 'contrivances'.

Defending Bennett's conceptualist reading, Timo Kajamies explains that the conceptualist analysis takes 'possibility' to mean the *same thing* as

¹There are some scholars who believe the cogito is *not* an argument and so would not contain any premises. See Jaakko Hintikka, 'Cogito Ergo Sum: Inference or Performance?' *Philosophical Review* 71 (1962): 3-32, and Hintikka, 'Cogito Ergo Sum as an Inference and a Performance' *Philosophical Review* 72 (1963): 487-496.

²Bennett (1994), 648.

‘conceivability’.¹ Our conceiving that some state of affairs *X* is possible does not refer to some possible state of affairs in the world, though our conception may represent this to us as being so. Kajamies argues that Bennett’s position on modal expressions can be captured by the following three definitions:

M1: *P* is impossible =_{df} for all subjects *s*: if *s* has *P* clearly and distinctly in mind, *s* is psychologically unable to assent to *P*.

M2: *P* is necessary =_{df} for all subjects *s*: if *s* has *P* clearly and distinctly in mind, *s* is psychologically unable to dissent to *P*.

M3: *P* is possible =_{df} for all subjects *s*: if *s* has *P* clearly and distinctly in mind, *s* is psychologically able to assent to *P*.²

From these definitions, we see how all modality is reducible to the language of conception for a subject and the appropriate judgment or attitude one takes toward some proposition. Modality, though it is established by God, signifies nothing more than a particular psychological state.

Both Bennett and Kajamies think this reading of modality strengthens the import of Descartes’s view in other passages that possibility and other modal notions are to be understood in terms of conceivability. In a letter to Arnauld, Descartes writes: ‘I merely say that [God] has given me such a mind that I cannot conceive a mountain without a valley, or an aggregate of one and two which is not three, and that such things involve a contradiction in my conception’ (AT V 224; CSMK 358-9). Instead of saying ‘it is impossible that there can be (or, God can make) a mountain without a valley’, Descartes knows he can only admit what is conceivable under a clear and distinct perception: ‘that such things involve a contradiction in my conception’. We can only assert that we cannot conceive a contradiction, but this is based on our inability to conceive it. Compare this to his letter to Gibieuf, 19 January 1642:

we cannot have any knowledge of things except by the ideas we conceive of them; and consequently, that we must not judge of them except in accordance with these ideas, and we must even think that whatever conflicts with these ideas is absolutely impossible and involves a contradiction. (AT III 476; CSMK 202)

Thus, we call something impossible or contradictory only because of the conflict of ideas in attempting to conceive them. Possibility, so the argument goes, is defined or understandable in terms of what is conceivable.

But are we really to believe that Descartes was himself a subjectivist about modality? Besides the Second Replies passage, what other textual evidence does the conceptualist have for attributing this view to him? Bennett believes

¹Kajamies (1999), 190.

²I find Kajamies’s reconstruction of Bennett’s definition of modal terms highly problematic. I postpone discussion of such problems for another forum.

that, broadly speaking, there is 'in Descartes a wide, deep, vivid streak of subjectivism or pragmatism about truth – a willingness to treat results about the settlement of belief as though they were results about how things stand in reality, or as though the former mattered and the latter did not'.¹ If we take Descartes's words to Gibieuf above seriously, especially his comment that we must never judge of things except in accordance with our ideas about them, this proposal by Bennett is forceful. Bennett's general claim is that both a subjectivist strand and an objectivist strand are present in the texts, and because both strands are there, either could be a candidate upon which to interpret Descartes's voluntarist thesis. Bennett chooses the subjectivist strand to interpret modality because he believes it 'harmonizes as well as anything can with the rest of Descartes's work'.² After all, the *Meditations* themselves are presented from a first-person point of view; what could be more subjective than that? Thus, the conceptualist interpretation of modality does not imply that Descartes was a thorough subjectivist with regard to all matters philosophical; as Bennett sees it, conceptualism about modality comports with other subjectivist strands as well as the objectivist strands in Descartes's works.

Problems with Conceptualism

While I think that Bennett's interpretation makes an admirable attempt to resolve the tension implicit in the creation doctrine, I find the view highly problematic. The most immediate problem concerns Bennett's principal supporting passage from the Second Replies. Despite his translation of the Latin *fingerere* as 'inventing', the passage does *not* entail that a possibility inhering in the object itself is strictly an invention or contrivance. It initially appears that Descartes presents us with a dilemma in this passage: either 'possibility' means 'whatever does not conflict with our human concepts', or it means something invented and hence not a real option. Bennett understands Descartes to be admitting the former and rejecting the latter. On closer inspection, however, this passage only claims that objective possibility is a contrivance when it fails to 'match' the first sort of possibility. There is no sweeping rejection of objective possibility here. Descartes tells the authors of the Second Objections that if they are talking about an objective possibility that is thoroughly divorced from a subjective understanding of that possibility (via a human concept), then that kind of possibility will be meaningless, for it would be unknowable. In this sense, objective possibility is a contrivance. But this does not entail the stronger claim that Bennett wants, namely, that *all* objective possibilities are contrivances. In this passage, Descartes allows that objective possibility may 'match' with a subjective understanding of that possibility, and in such cases, one may speak coherently of an objective possibility. There is no easy way to understand what Descartes means by an objective possibility 'matching' a subjective understanding of possibility, but I

¹Bennett (1994), 652.

²*Ibid.*

do not think that 'equals' is the right way to render it, as Bennett seems to be doing.

One way we might construe this matching of the subjective to the objective is through Descartes's remarks to an unknown correspondent, regarding the foundations of our thoughts. Though the letter discusses the theory of distinctions, Descartes makes a comment which puts importance on distinguishing our thoughts from the objects of those thoughts. He writes:

I do not recognize any distinction made by reason *ratiocinantis* – that is, one which has no foundation in reality – because we cannot have any thought without a foundation; ... It seems to me that the only thing which causes difficulty in this area is the fact that we do not sufficiently distinguish between things existing outside our thought and the ideas of things, which are in our thought. (AT IV 349-50; CSMK 280)

Descartes admonishes us to always distinguish carefully between our ideas of things and the existence of those things in reality, that is, outside our thoughts about them. Why is this important? In the context of this letter, one should distinguish carefully so as to avoid postulating non-existent entities of which one may (mistakenly) think one has a clear perception. Now, this passage does not mention possibility or any modal terms, so there is no direct support for the Second Replies claim. However, since Descartes is speaking generally about thoughts being grounded in an external reality, this would apply to any thoughts about possible or necessary states of affairs. The main point is that there must be *some* foundation for the thought, else Descartes does not recognize that we can even have such thoughts.

An even stronger passage conflicting with Bennett's view can be found in the 30 September, 1640 letter to Mersenne. Descartes accepts a general principle about conceivability and possibility, yet the language would seem to indicate that he allows for both a subjective *and* objective understanding of possibility and how they match up:

I entirely agree with the argument... that whatever we conceive distinctly to be possible is possible, and that we conceive distinctly that it is possible that the world has been made, and therefore it has been made. (AT III 191; CSMK 154)

The subjective reading is contained in the phrasing 'whatever we conceive distinctly to be possible', and the objective reading is captured by 'is possible'. The principle holds that when we conceive the matter to be possible (subjectively), it *is* possible (objectively). It seems apparent that Descartes intends two kinds of possibility here, yet it is not clear that one is being reduced to or equated with the other. Rather, conceiving as possible is an *indication* of or guide to the objective possibility. Of course, the way we come to know possibilities is through our concept of possible, which as indicated

above, is to conceive the matter as being possible. So we might take this passage to support the one in the Second Replies such that there must be a proper match between possibility-as-we-conceive-it and possibility-as-it-is for the latter to have any meaning.

The example of God's necessary existence may be the biggest difficulty for the proponent of conceptualism. In the Fifth Meditation, Descartes clearly states that God is the only being whose essence is *necessary* existence: 'apart from God, there is nothing else of which I am capable of thinking such that existence belongs to its essence. ... For what is more self-evident than the fact that the supreme being exists, or that God, to whose essence alone existence belongs, exists?' (CSM II 47; AT VII 68-69). Admittedly, this passage sounds subjective, for Descartes implies that God's essence is existence because he conceives of it self-evidently. On a closer reading, it is apparent that these claims are not *argued* subjectively, even though they are expressed subjectively. First, Descartes says that there is no other being or object he can think of such that its essence is to exist. He is not claiming that his thinking this way makes it the case that God is the only being whose essence is to exist. Consider this earlier passage in the same Meditation:

But from the fact that I cannot think of God except as existing, it follows that existence is inseparable from God, and hence that he really exists. It is not that my thought makes it so, or imposes any necessity on any thing; on the contrary it is the necessity of the thing itself, namely the existence of God, which determines my thinking in this respect. (CSM II 46; AT VII 67)

It is clear that God's necessary existence is what determines Descartes to think in this way about God. To be sure, Descartes begins with a subjective claim: his conception that God cannot but exist. But the conclusion does not refer to his thought; it is a metaphysical claim. But what is more devastating for the conceptualist is Descartes's explicit claim that his thought does not impose necessity *on any thing*, especially the necessity of God's existence. It is the objective or mind-independent necessity which determines one's thought, and thus determines or constitutes one's clear and distinct perception of the matter. Bennett comes close to addressing this objection.¹ He raises this idea about God's existence, but his cashing out the proposition, 'Necessarily God exists', as something about our conceptions seems to miss what Descartes says about God's essence. On Bennett's view, this statement might read: 'we can't but conceive God as existing', but this seems too limiting. Descartes makes the stronger claim that God necessarily exists despite what we might believe. After all, Descartes surely holds that God exists despite what the atheist believes. Bennett is right that Descartes often put propositions about what is really the case in terms of what to believe, but there is little evidence that Descartes equates them or reduces one to the other.

¹Bennett (1994), 649-650.

Conclusion

I have argued that the conceptualist analysis of modality ultimately cannot constitute Descartes's theory. Unlike Bennett, I believe that Descartes's modal theory should be informed by what he argues across his writings. One important area is Descartes's physics, where he reportedly first came upon the creation doctrine. Descartes argues that the laws of nature, which Descartes classifies as eternal truths, are created by God as metaphysical necessities inhering in the world. This would seem to go beyond their creation as merely conceptual entities. It seems that Descartes must have conceived eternal truths as (1) separate from, but not independent of, God; and (2) separate from, and not dependent on, human minds; thus, Descartes seems committed to a Platonic view about these truths. In his work on physics, *Le Monde*, Descartes claims that the three principal laws of nature follow from certain principles about God. He also calls these principles eternal truths. These remarks are repeated in his seminal philosophical text, the *Principles of Philosophy*. And given the early letter to Mersenne from 15 April 1630, Descartes clearly states that 'it is God who has laid down these laws in nature just as a king lays down his laws in his kingdom'.¹ It would be quite an interpretive stretch to read 'in nature' here as 'in human minds' if we were convinced by the conceptualist argument. The laws of nature Descartes describes turn out to be mathematical truths, which are eternally true.² Thus, we see how Descartes conceived God's decrees to form the foundation of his physics, decrees which are independent of, but certainly knowable by, human minds.

The guiding idea in these and other passages points to a unique aspect of modality for Descartes: it is a creation of God that inheres in the world, and so is *not* strictly a subjective feature of human minds. To be sure, Descartes is *not* rejecting a metaphysical account of modality; he is merely saying that a metaphysical account of modality divorced from our conceptual understanding of it will be unwarranted. The troubling aspect is that Descartes speaks in the subjective or conceptual language in discussing possibility and necessity, and many commentators have interpreted this conceptual language to be a clear indication of Descartes's settled conceptualist view regarding modality. I find Descartes's use of conceptual language in discussing modality troubling insofar as it does not neatly distinguish between real possibilities and necessities on the one hand, and conceived possibilities and necessities on the

¹Cf. *Discourse on Method*, Part Five: 'I have noticed certain laws which God has so established in nature, and of which he has implanted such notions in our minds, that after adequate reflection we cannot doubt that they are exactly observed in everything which exists or occurs in the world' (AT VI 41; CSM I 131).

²One problem facing the conceptualist interpretation is how something eternal can have its ontological seat in something finite like a human mind. This problem has motivated some, like Marleen Rozemond, to argue instead for a neo-Platonic conception of eternal truths, whereby the ontological status of eternal truths is God's mind. See her 'Descartes's Ontology of the Eternal Truths' in *Contemporary Perspectives on Early Modern Philosophy: Essays in Honor of Vere Chappell*, eds. Paul Hoffman, David Owen, and Gideon Yaffe (Peterborough, Ontario: Broadview Press, 2008), 41-63.

other. However, as I have shown above, there are several texts which support the interpretation that Descartes does *not* endorse a conceptual view of modality, despite the conceptualist-sounding language. Even so, I believe Descartes uses conceptual language simply because he was trying to convey the certainty that our concepts hook up to the world in the way we conceive them to when we are clear and distinct about them.

References

- Adam, Charles & Paul Tannery, eds. (1897-1913) *Oeuvres de Descartes*, 11 volumes. Paris: Léopold Cerf.
- Alanen, Lilli. (1991) 'Descartes, Conceivability, and Logical Modality.' In: T. Horowitz & G. J. Massey (eds.), *Thought Experiments in Science and Philosophy*, 65-84. Savage, Maryland: Rowman & Littlefield.
- Bennett, Jonathan. (1994) 'Descartes' Theory of Modality.' *Philosophical Review* 103: 639-667
- Bouveresse, Jacques. (1983) 'La théorie du possible chez Descartes.' *Revue Internationale de Philosophie* 146: 293-310.
- Cottingham, J., R. Stoothoff, & D. Murdoch, eds. (1984-1985) *The Philosophical Writings of Descartes*, Volumes I & II. Cambridge: Cambridge University Press.
- Cottingham, J., R. Stoothoff, D. Murdoch, & A. Kenny (1991) *The Philosophical Writings of Descartes*, Volume III. Cambridge: Cambridge: University Press.
- Hintikka, Jaakko. (1962) 'Cogito Ergo Sum: Inference or Performance?' *Philosophical Review* 71: 3-32.
- Hintikka, Jaakko. (1963) 'Cogito Ergo Sum as an Inference and a Performance' *Philosophical Review* 72: 487-496.
- Ishiguro, Hidé. (1986) 'The Status of Necessity and Impossibility in Descartes.' In: A. Rorty (ed.), *Essays on Descartes's Meditations*, 459-471. Berkeley: University of California Press.
- Kajamies, Timo. (1999) 'The Concept of Power and the Eternity of the Eternal Truths in Descartes.' *Southern Journal of Philosophy* 37: 189-200.
- Rozmond, Marleen. (2008) 'Descartes's Ontology of the Eternal Truths.' In: P. Hoffman, D. Owen, & G. Yaffe (eds.), *Contemporary Perspectives on Early Modern Philosophy: Essays in Honor of Vere Chappell*, 41-63. Peterborough, Ontario: Broadview Press.
- Wilson, Margaret. (1978) *Descartes*. London: Routledge & Kegan Paul.

CHAPTER ELEVEN

Hegel and Foucault Re-visited

Evangelia Sembou

It is commonly assumed that Hegel and Foucault are so different thinkers that they are beyond comparison. Hegel is one of the foremost modernist philosophers, a systematic thinker, associated with teleological grand narratives. By contrast, Foucault is a postmodernist, post-structuralist thinker and a deconstructionist. Foucault himself regarded Hegel's philosophy as a closed system and criticized Hegelianism, although his remarks are often quite ambiguous (e.g. 1966: 318, 338, 339, 342, 345/2002a: 335, 356, 357, 361, 364; 1971a: 74-79/1981a: 74-76; 1977a: 145/1980c: 114-115; 1980d: 43, 48, 49-50/2002b: 241, 246-247, 248, 249; 2001: 29-30, 466-467/2005: 28, 477-489).

I will suggest that, their differences notwithstanding, Hegelian 'phenomenology' and Foucauldian 'genealogy' make a common contribution to philosophy by questioning the 'correspondence theory of truth', that is, the theory that asserts that truth consists in a correspondence between thought (the subject) and the objective world. My claim is not that the central concern of Hegel and Foucault is to overthrow the 'correspondence theory'; rather, my aim is to show that Hegel's 'phenomenology' and Foucault's 'genealogy' can bring important insights to the central theme of philosophy, namely, truth. Hegel and Foucault had a lot to say about truth, although no reference is made to either of them in discussions on (the nature of) truth carried out by analytical philosophers. The purpose of this paper is, therefore, to make a substantive contribution to discussions on truth by bridging the gap between 'analytical' and 'continental' philosophy, so as to enable a fruitful dialogue.

Moreover, I will suggest that the aforesaid approaches share a non-foundational conception of knowledge.¹ My reading of Hegel falls into that tradition of Hegel scholarship that advances a non-metaphysical and non-foundational understanding of Hegel (e.g. Hartmann, 1972; Westphal, 1979; Rose, 1981; Houlgate, 1986 & 2004; Maker, 1994; Sallis, 1995; Hutchings, 2003).

Hegel's 'Phenomenology'

Hegel defines his *Phenomenology of Spirit* as 'Science of the *experience of consciousness*' (*PhG*: 74/*PhS*: 56). 'Consciousness' is internally divided into two aspects ('moments'), the knowing subject and the object (*PhG*: 32/*PhS*: 21).

¹All citations will be from the relevant English translation, unless otherwise stated.

‘Consciousness’ develops into ‘spirit’, and the *Phenomenology* consists in an exposition of this development. Because the *Phenomenology* presents this development as *necessary*, it is ‘Science’ (*Wissenschaft*) (*PhG*: 74/*PhS*: 56).

The *Phenomenology* challenges what underlies the ‘correspondence theory’, namely, the idea that there is a fundamental distinction between the subject and the objective world, by showing that consciousness is the dynamic cognitive interrelationship between the knowing subject and the known object (*PhG*: 70/*PhS*: 52). Therefore, there is no need of an external criterion to determine the validity of consciousness’s knowledge at each stage of its development. Consciousness attempts to make its subjective and objective aspects correspond to each other (*PhG*: 71/*PhS*: 53). It in turn takes one of its aspects to be necessary (or stable) and measures the validity of its other aspect against it. Thus, there is nothing for the philosopher to do but ‘simply to look on’. In the course of this self-examination, consciousness is both ‘consciousness of the object’ and ‘consciousness of itself’. Once consciousness realizes that one of its ‘moments’ (the subject) does not correspond to the other (the object), it alters it. But, following the change of the subjective aspect (‘knowledge’), the object changes as well, for the knowledge in question was ‘a knowledge of the object’. Consequently, what consciousness took to be the object or ‘the *in-itself*’ turns out to be ‘an *in-itself for consciousness*’. Consciousness now takes another object to be the truth and, insofar as this emerges from within consciousness itself, consciousness undergoes a ‘*dialectical* movement’. This constitutes its ‘*experience*’ (*Erfahrung*) (*PhG*: 72-73/*PhS*: 54-55).

From a phenomenological point of view, it is irrelevant whether there are any objects outside consciousness’s experience; what matters is the way(s) consciousness cognitively interacts with the world. Consequently, ‘the True’ is not an independent entity (an ‘*in-itself*’) but ‘the being-for-consciousness of this *in-itself*’ (*PhG*: 73/*PhS*: 55). Phenomenological experience is distinctive in that the knowing subject’s realization of the untruth of its first object does not come about once the subject runs across another object by chance; nor does the second object come externally, as it were. Rather, the new object is the result of ‘a *reversal of consciousness itself*’. Without intervening with the experience of consciousness, the phenomenological observer (the ‘we’) elevates consciousness’s successive reversals ‘into a scientific progression’ (*PhG*: 73-74/*PhS*: 55), as he is able to see that the negation of one object does not result in nothingness but in a new object.¹ This occurs ‘behind the back of consciousness’, however (*PhG*: 74/*PhS*: 56). By perceiving the consecutive reversals of consciousness and presenting them in a phenomenological account, the philosopher shows their *necessity*. Within this phenomenological exposition the different ways in which consciousness attempts to grasp reality appear as ‘*patterns of consciousness*’ (*Gestalten des Bewußtseins*) (*PhG*: 74-75/*PhS*: 56).

‘Phenomenology’ takes the object of knowledge to be an aspect of cognition itself; so the acquisition of knowledge and the development of the object are closely associated. In the course of the *Phenomenology* both the

¹*PhG*: 68-69/*PhS*: 50-51 (‘determinate negation’).

knowing subject and the object of knowledge become more and more comprehensive. Initially, the subject appears as consciousness ('sense-certainty', 'perception', 'understanding') and the object first as external reality (a mere 'This' in the form of 'Now' and 'Here', then a 'Thing') and subsequently as consciousness (the notion of force, the 'realm of laws', 'infinity'). As soon as consciousness realizes that its conception of the world as external is inadequate, it reconsiders its interaction with the world. Now the subject appears as 'desire' and the object as life in its totality. But the subject as 'desire' cannot reach satisfaction; once it destroys (consumes) one object, it is overcome by a new desire and this process goes on forever. The subject becomes a self-consciousness and its object (another) self-consciousness (life-and-death struggle, master-slave relation, stoicism, scepticism, the unhappy consciousness). Self-consciousness reaches its limits in the experience of the 'unhappy consciousness', which cannot reconcile its changeable and unchangeable aspects and thereby collapses; out of it emerges another form of understanding, namely, 'Reason'. Now self-consciousness is certain that it is all reality (idealism); reason permeates reality, it has both a subjective and an objective aspect. Following its failure (as 'observing Reason') to adequately comprehend its object (it foolishly grasps the essence of the self to be the skull) (*PhG*: 240ff/*PhS*: 197ff), the subject gives up its attempt to *find* itself in objects and tries to *impose* itself on the objective world through action (the hedonist, the romantic and the quixotic self-consciousness; the 'spiritual animal kingdom'; 'reason as lawgiver' and 'reason as testing laws'). Once 'Reason' proves to be untenable on its own terms (anything can pass the universalizability test inasmuch as it is formally self-consistent), a new form of understanding emerges, namely, 'Spirit'. Each 'shape' of spirit has both a subjective and an objective aspect, and an immanent development is brought about by the failures of all forms of life (Greek antiquity, the Roman Empire, feudal Europe, pre-revolutionary and revolutionary France, the moral world-view and the community of 'beautiful souls') to attain an adequate understanding of the principles underlying their existence. In 'Religion' the subject is self-conscious spirit and the object is absolute spirit. The religious form of understanding develops from 'Natural Religion' through the 'Religion of Art' to the 'Revealed Religion' as self-conscious spirit attempts to attain an awareness of itself. Although in the 'Revealed Religion' self-conscious spirit comes to grasp its essence, its self-understanding is imperfect due to the form of its knowledge ('*Vorstellung*') (*PhG*: 480/*PhS*: 416).¹

'Absolute knowing'² is the realization that all attempts by humans to grasp reality in terms of the 'correspondence theory of truth' have failed on their own terms (*immanent critique*).³ To say that Hegel's 'phenomenology' is an

¹Miller translates '*Vorstellung*' as 'picture-thought'.

²It is *knowing* (*Wissen*), not *knowledge* (*Kenntnis*); this implies that it is an approach or stance. In Hegelian terminology 'absolute' is something that is not conditioned by anything else, hence *self-determining*.

³Therefore, I disagree with Harris (1997) and Westphal (1997) when they say that Hegel has a 'correspondence theory of truth'. Harris sees a 'correspondence theory of truth' of a special

immanent critique of the ‘correspondence theory’ is not the same as to say that its purpose is to abolish the subject/object distinction. This interpretation sees ‘absolute knowing’ as something positive, to wit, an adequate, comprehensive form of knowledge.¹ Rather, I mean that its purpose is negative;² that is, to demonstrate what philosophical thinking is *not*. Hegel’s ‘phenomenology’ shows that knowledge is conditioned by spirit, as all the ‘shapes of consciousness’ prior to ‘Spirit’ presuppose spirit, which is intersubjective and self-determining, defined as “‘I’ that is “‘We” and “‘We” that is “‘I”” (*PhG*: 140/*PhS*: 110).³

Foucauldian ‘Genealogy’

Foucault’s Nietzschean genealogy attacks the notion of ‘origins’. Nietzsche had used the terms ‘*Ursprung*’, ‘*Entstehung*’ and ‘*Herkunft*’ mostly interchangeably (*NGH* (F): 137-138/*NGH* (E): 77-78). For Foucault, genealogy opposes ‘*Ursprung*’ in particular. First, ‘*Ursprung*’ suggests “‘that which was already there””, a primeval truth. Nevertheless, genealogy uncovers that the essence of things is a product of history and that the notion of the ‘origin’ is one of those metaphysical ideas that have dominated philosophy. The idea of the origin implies the ‘correspondence theory’ because it presumes that the origin is some ‘fact’ to be discovered; truth would then consist in a relation of correspondence between a proposition and this ‘fact’. However, groping into the ‘history of reason’, the genealogist discovers that it emerged from chance, from the competition among scholars (*NGH* (F): 138/*NGH* (E): 78), what Nietzsche called the ‘will to truth’ (1976b: 30-31/1980a: 66). Second, ‘*Ursprung*’ connotes that the beginning is always the moment of perfection. As such, it appears as an objective reality, a ‘fact’ to be found. However, genealogy uncovers that ‘historical beginnings are lowly’. Third, ‘*Ursprung*’ implies ‘the site of truth’, ‘...the point where the truth of things corresponded to a truthful discourse...’ (*NGH* (F): 139/*NGH* (E): 79). But genealogy demonstrates that faith in an original truth subsequently veiled by historical development is just a metaphysical illusion. Actually, the very idea of ‘truth’ is an error (*NGH* (F): 139-140/*NGH* (E): 79-80).

Genealogy concerns itself with ‘*Herkunft*’ and ‘*Entstehung*’. A genealogical analysis of ‘*Herkunft*’ (descent) is an exercise of deconstruction; it fragments what were considered to be unitary entities and decomposes ideas into their constituent elements. Genealogy concentrates on the accidental, the event (*NGH* (F): 140-141/*NGH* (E): 80-81); and it shows that there is no given reality to which language could correspond. Moreover, genealogy explores the emergence

kind in Hegel’s *Phenomenology of Spirit* and not what is ordinarily understood by ‘correspondence theory of truth’ (p. 11).

¹Most commentators support this view; e.g. Bloch (1951/1949), Findlay (1958: 144-148), Taylor (1993/1977: 48-49, 119, 214-221), Hyppolite (1946), Kojève (1947), Gadamer (1971), Pöggeler (1973), Heinrichs (1974), Fulda (1975), Lauer (1976), Westphal (1979).

²I am following Bubner (1970) and Maker (1994: Part One, esp. 71-82, 86-89, 89-93, 100-106).

³My reading here owes much to Rose (1981) and Hutchings (2003: 32-44).

(‘*Entstehung*’) of a battle which determines a space (NGH (F): 143-144/NGH (E): 83-84); this latter is ‘a “non-place,” a pure distance, which indicates that the adversaries do not belong to a common space’. Consequently, emergence ‘always occurs in the interstice’ (NGH (F): 144/NGH (E): 85).

Genealogy also challenges the ‘correspondence theory of truth’ by denying there is ‘something’ to be interpreted. It uncovers that the history of humankind is but a sequence of interpretations and that each new interpretation is contingent on some newly emergent configuration of power (NGH (F): 145-146/NGH (E): 86; cf. Foucault, 1977a: 145/1980c: 114). The dominance of the ‘correspondence theory’ has largely been due to the commonsense belief that ‘Whether what is said about the world is true surely must depend on how the world is’ (Rundle, 2005: 178). But genealogy’s refusal to acknowledge the existence of a substratum of reality strikes a blow, as it were, to any attempt to establish a supposedly exact correspondence between thought and objective reality.¹ And, if in his early work Foucault had presupposed a primordial essence of madness (1961; 1995/1971), by the 1970s he had come to see that there was no substantial reality to be grasped.

Additionally, genealogy questions the very notions which the ‘correspondence theory’ presupposes, namely, the subject and the object. It explores how the constitution of objects (e.g. madness or criminality) takes place within history. Simultaneously, it queries the idea of the ‘constituent subject’ and analyzes ‘the constitution of the subject within a historical framework’ (Foucault, 1977a: 147/1980c: 117; ‘dé-subjectivation’/‘desubjectivation’, 1980d: 43/2002b: 241). Genealogy shows that the human body too is dissected, ‘cut into pieces’ (NGH (F): 147/NGH (E): 87). Challenging essentialist notions of the self like the Cartesian ego and the Kantian self, genealogy demonstrates that the subject is constructed (Foucault, 1982). The subject is constructed by investing a body with certain habits; it is then a subject in the double sense of being subject to disciplinary mechanisms and of being a subject of experience (SP/DP). Moreover, power manufactures subjects as sexual beings; concurrently, this understanding of subjectivity is implanted in people’s minds (HSI (F)/HSI (E)).

Challenging traditional history, genealogy as ‘effective history’ queries the assumption that there are ‘facts’ to be interpreted; rather, ‘facts’ themselves are constructions. As opposed to the ‘correspondence theory’ which asserts that there is *only one* correct relation of correspondence of thought to objective reality, viz. *one truth*, genealogy reveals that there is *no single* truth but truths instead, these truths being no more than interpretations. For example, *Discipline and Punish* shows that from the Middle Ages until the great penal reforms of the eighteenth and nineteenth centuries ‘to judge was to establish the truth of a crime...’; one needed to have ‘Knowledge of the offence, knowledge of the offender, knowledge of the law...’ (SP: 26/DP: 19). However, following the penal reforms, the question is no longer simply whether a crime has been committed, by whom and on the basis of what law the perpetrator should be

¹The best (and, perhaps, only) study that considers Foucault’s challenge to the ‘correspondence theory of truth’ is Prado (2000).

punished, but ““What is this act...? [...] Is it a phantasy, a psychotic reaction, a delusional episode, a perverse action?””, what were the perpetrator’s motives and how can the offender be rehabilitated? (*SP: 27/DP: 19*; Foucault refers to a ‘scientifico-judicial complex’.) According to article 64 of the 1810 Code, there was no offence if the perpetrator was mentally ill. Gradually the judges came to interpret this as stating that the gravity of the offence should be determined according to the degree of sanity or insanity of the malefactor (*SP: 27-28/DP: 19-21*). Consequently, nowadays psychiatrists, psychologists, educationalists, officials who implement the sentences and prison officers are involved in ‘the administration of the penalty’ (*SP: 29/DP: 21*); psychiatry is called to advise on the criminal’s ‘medico-judicial treatment’ (*SP: 29/DP: 22*). So genealogy uncovers that what underlies the decreasing severity of punishment is ‘a whole new system of truth’, which ‘becomes entangled with the practice of the power to punish’ (*SP: 30/DP: 23*).

To mention another example: In *The History of Sexuality*, Volume 1, Foucault asks in respect of what he calls the ‘repressive hypothesis’: ‘Is sexual repression truly an established historical fact [*une évidence historique*]?’ (*HSI (F): 18/HSI (E): 10*) The explosion of discussion about sex in the Victorian age was due to a ‘type of power’ which bourgeois society ‘brought to bear on the body and on sex’. A genealogical study shows that this power operated by producing (different kinds of) sexuality and making it a defining characteristic of individuals (*HSI (F): 64-65/HSI (E): 47*). So, far from being an historical fact, sexuality is ‘a historical construct’ (*«un dispositif historique»*) (*HSI (F): 139/HSI (E): 105*). Therefore, the real questions are whether prohibition and censorship are not forms of power rather than repression and whether all this discourse on sex is not itself part of the power it criticizes as ‘repression’ (*HSI (F): 18/HSI (E): 10*).

In sum, genealogy unmask the politics of truth. Foucault usually refers to ‘régimes of truth’ (*«régimes de vérité»*) (1977a: 143, 158, 160/1980c: 112, 131, 133).¹ Truth/knowledge and power (relations) implicate each other (Foucault, 1977a: 158-160/1980c: 131-133; 1977c: 175-176/1980b: 93; *SP: 36/DP: 27*; *HSI (F): 80-81/HSI (E): 60*),² hence Foucault’s term ‘power-knowledge’ (*«pouvoir-savoir»*) (*SP: 36/DP: 27-28*). However, truth and knowledge are not reducible to power. For Foucault, the attempt to think of truth/knowledge as *either* distinct from (opposed to) power *or* as determined by power is to yield to ‘the intellectual and political blackmail of “being for or against the *Enlightenment*”’ (1984c: 573/1984d: 45).³ Actually, Foucault insisted that power presupposes resistance and vice versa (*HSI (F): 125-127/HSI (E): 95-96*).⁴

¹For a ‘régime of truth’ versus ideology see Foucault (1977b: 183-184/1980b: 102). For genealogy versus a critique of ideology see Owen (2002).

²See also Foucault (1971a: 10-47/1981a: 52-64).

³My translation.

⁴Cf. Foucault (1982: 211-212, 221-222, 225-226).

Toward a Non-Foundational Knowledge

The ‘correspondence theory’ is typical of foundationalism. Foundationalism starts from the assumption that there is a fundamental distinction between the subject of knowledge (the Cartesian ego) and the objective world. It follows that there must be some standard independent of knowledge whereby this knowledge can be tested. The test consists in determining whether there is a correspondence between knowledge and the object. Therefore, foundationalism is simultaneously trying to do two things: to retain the distinction between knowledge and the object, so that the comparison between knowledge and the object can be carried out; and to show that knowledge and the object coincide. Yet this is impossible.¹

According to the ‘correspondence theory’, knowledge is descriptive; that is, knowledge is a mirror of objective reality. However, both Hegelian phenomenology and Foucauldian genealogy question this conception of cognition; Hegel’s phenomenology by showing that conscious experience is a dynamic interrelationship of subject and object; Foucault’s genealogy by demonstrating that both the subject and the various objects of knowledge are constituted in discourses. By questioning the idea of an objective reality apart from human experience and activity, both approaches query the notion of the ‘given’. They thus challenge one more feature of foundationalism, namely, the view that all knowledge rests on (certain) presuppositions. They also query the ground of foundationalist thinking, what May has called ‘the space of interiority’ (1993: 57). This latter is implicitly questioned in the genealogy of psychology; for what psychological discourse assumes is the existence of the mind, whose structure is conceived in ahistorical and transcendental terms. It is on the human mind that traditional philosophizing has found its foundations (May, 1993: 57-59). ‘Subjective foundationalism’² or subjectivism assumes that the objective world lies out there separate from subjectivity, while the subject provides the foundations of knowledge. However, as shown above, Hegelian phenomenology consists in an immanent critique of the type of knowing that assumes that there is a distinction between the knowing subject and the object known.

In saying that for Hegel and Foucault there is no objective reality apart from human experience and activity, I do not mean that Hegel and Foucault denied the existence of the external world. In the case of Hegel, this would mean that the objective world is a product of the human mind (extreme idealism); in the case of Foucault, it would imply that Foucault denies the existence of the world (irrealism).³ Both these positions are wrong. What I mean is that, from a Hegelian-phenomenological and a Foucauldian-genealogical standpoint, it is irrelevant what the world apart from human experience is like. What is relevant

¹See Maker (1994: 60-63). I am partially following Maker.

²May (1993: 59).

³For an argument against the view that Foucault’s understanding of truth leads to irrealism see Prado (2000: Chapter 7).

is how humans interact with this world, how they conceptualize it and how they make sense of it.¹

The question of non-foundationalism is an intricate one. What troubles critics is how such a non-foundational knowledge can justify itself.² As regards Hegelian phenomenology, it grounds itself internally. Initially, science comes on the scene; it *appears*. As such, it cannot justify itself against ‘another mode of knowledge’ (*PhG*: 66/*PhS*: 48). The appearance of science (i.e. phenomenology) consists in its identifying step by step what it itself is. Hegelian phenomenology is the immanent development of foundationalist thinking and its model of knowledge. Foundationalist thinking collapses internally, as all attempts to ground knowledge in terms of the ‘correspondence theory of truth’ fail by their own standards. Therefore, Hegelian phenomenology shows what philosophical science is not like. It is in this way that the *Phenomenology* is an introduction to the *Science of Logic*. But ‘to know this is not a presupposition, simply because this negative knowledge is not necessary for science’ (Maker, 1994: 93). As to Foucauldian genealogy, a useful distinction is that drawn by May between ‘grounds’ and ‘foundations’ (1993: 11). A ‘ground’ has to do with the way one justifies one’s claims, whereas a ‘foundation’ is some ultimate truth that cannot be doubted. The ultimate truths foundationalism takes for granted are, to name just a few, the subject of knowledge, the mind and the soul. But these are notions that Foucault’s genealogies question. The issue, then, is to see how Foucault’s genealogies justify themselves, without using these or any other truths as foundations (May, 1993: 67, 71-72). However, May’s distinction between ‘justification’ and ‘truth in an ultimate sense’ is problematical (1993: 71; cf. 93). May’s reference to ‘truth in an ultimate sense’ is misleading, since this is precisely what Foucault’s genealogy calls into question.³ Considering his distinction between ‘grounds’ and ‘foundations’, May (1993) is right to point out two aspects of justification: ‘the inferential move itself and the status of the claim to which the inferential move appeals in its attempt at justification.’ The former is logical (deduction, induction); it is the latter that differentiates Foucault’s understanding of knowledge from foundationalism (p. 90). From a Foucauldian-genealogical perspective, the truth of a claim is contingent on the place it occupies and on its role within a specific discourse. Genealogy acknowledges its interested character (perspectivism) (*NGH* (F): 150/*NGH* (E): 90). Its significance lies in that it provides an alternative picture or interpretation. By practising ‘local’ criticism, genealogy allows ‘an *insurrection of subjugated knowledges*’ (1977b: 163/1980b: 81). Foucault distinguished between the ‘universal’ and the ‘specific’ intellectual; whereas the former establishes universal norms, the latter provides specific analyses (1977a: 159/1980c: 132). Ultimately, what vindicates genealogy is its documentary character and attention to detail (*NGH* (F): 136/*NGH* (E): 76-77; Foucault, 1980d: 44/2002b: 242). A

¹With reference to Foucault, cf. Prado (2000: 147-148).

²Respecting Foucauldian genealogy, the question of grounds has been raised by Fraser (1981), Dews (1988/1987: Chapters 5-7, esp. 161-170, 173-176, 180-185, 186-192, 192-199, 214-216, 218-220) and Habermas (1985: Vorlesungen IX u. X/1990: Lectures IX and X).

³Cf. Prado, 2000: 140.

genealogical account can be rejected if, and *only* if, another historical account (a genealogy of genealogy) can prove to be more convincing.¹ In fact, perhaps it is wrong to expect genealogy to justify itself in advance. Given that its purpose is to disabuse us of the foundations that have underpinned traditional philosophy, its value can be judged only retrospectively;² its worth can be proven after it has successfully (or otherwise) enabled us to change our way of thinking (1980d: 44/2002b: 242). Therefore, both Hegelian phenomenology and Foucauldian genealogy can be assessed, as well as justify themselves, only in retrospect.

Conclusion

I have argued that, despite their many differences, Hegelian phenomenology and Foucauldian genealogy pose a common challenge to philosophy by questioning the ‘correspondence theory of truth’. By casting doubt on the ‘correspondence theory’, they also query foundationalism and point towards a non-foundational knowledge. Both approaches demonstrate that truth is open-ended; truth and knowledge are intertwined with human activity and social-cum-political life. This implies that the philosopher, too, is a product of his time and, hence, his philosophy is part of the given social and political context in which he finds himself.³

References

- Bloch, E. (1951, c.1949). *Subjekt-Objekt: Erläuterungen zu Hegel*. Berlin: Aufbau-Verlag.
- Bubner, R. (1970). ‘La Philosophie et ses Apparences.’ *Actes du IIIème Congrès International de l’ Association pour l’ Étude de la Philosophie de Hegel*. Lille: Association des Publications de la Faculté des Lettres et Sciences Humaines, 65-81.
- Dews, P. (1988, c.1987). *Logics of Disintegration: Poststructuralist Thought and the Claims of Critical Theory*. London and New York: Verso.
- Findlay, J. N. (1958). *Hegel: A Re-examination*. London: George Allen & Unwin; New York: Macmillan.
- Foucault, M. (1961). *Folie et Dérison: Histoire de la Folie à l’Age Classique*. Paris: Plon.
- Foucault, M. (1966). *Les Mots et les Choses: Une Archéologie des Sciences Humaines*. Paris: Gallimard.
- Foucault, M. (1995, c.1971). *Madness and Civilization: A History of Insanity in the Age of Reason*. Trans. R. Howard. London: Routledge.
- Foucault, M. (1971a). *L’ Ordre du Discours*. Paris: Gallimard.
- Foucault, M. (1971b). ‘Nietzsche, la Généalogie, l’ Histoire’ [NGH (F)]. In D. Defert *et al* (éds.) (1994), *Dits et Écrits, 1954-1988*, Tome II, 136-156. Paris: Gallimard.

¹For a similar, albeit slightly different, argument see May (1993: 100-101).

²Cf. Prado (2000: 175).

³For the politics of Hegel’s and Foucault’s philosophy see Hutchings (1999: 93-117).

- Foucault, M. (1975). *Surveiller et Punir: Naissance de la Prison* [SP]. Paris: Gallimard.
- Foucault, M. (1976a). *Histoire de la Sexualité*. Vol. 1: *La Volonté de Savoir* [HSI (F)]. Paris: Gallimard.
- Foucault, M. (1976b). 'Questions à Michel Foucault sur la Géographie.' In D. Defert *et al* (éds.) (1994), *Dits et Écrits. 1954-1988*, Tome III, 28-40. Paris: Gallimard.
- Foucault, M. (1977a). 'Entretien Avec Michel Foucault.' In Defert *et al* (éds.) (1994), *Dits et Écrits*, III, 140-160.
- Foucault, M. (1977b). 'Cours du 7 Janvier 1976.' In Defert *et al* (éds.) (1994), *Dits et Écrits*, III, 160-174.
- Foucault, M. (1977c). 'Cours du 14 Janvier 1976.' In Defert *et al* (éds.) (1994), *Dits et Écrits*, III, 175-189.
- Foucault, M. (1980a). 'Questions on Geography.' In C. Gordon (ed.), *Power/Knowledge: Selected Interviews and Other Writings 1972-1977*, 63-77. Hemel Hempstead: Harvester Wheatsheaf.
- Foucault, M. (1980b). 'Two Lectures.' In Gordon (ed.), *Power/Knowledge*, 78-108.
- Foucault, M. (1980c). 'Truth and Power.' In Gordon (ed.), *Power/Knowledge*, 109-133;
- Foucault, M. (1980d). 'Entretien Avec Michel Foucault.' In D. Defert *et al* (éds.) (1994), *Dits et Écrits, 1954-1988*, Tome IV, 41-95. Paris: Gallimard.
- Foucault, M. (1981a). 'The Order of Discourse.' Trans. I. McLeod. In R. Young (ed.), *Untying the Text: A Post-Structuralist Reader*, 48-77. Boston, London, Henley: Routledge & Kegan Paul.
- Foucault, M. (1981b). 'Est-il Donc Important de Penser?' In Defert *et al* (éds.) (1994), *Dits et Écrits*, IV, 178-182.
- Foucault, M. (1982). 'The Subject and Power.' In H. L. Dreyfus & P. Rabinow, *Michel Foucault: Beyond Structuralism and Hermeneutics*, 208-216. Brighton: Harvester.
- Foucault, M. (1984a). 'Nietzsche, Genealogy, History' [NGH (E)]. In P. Rabinow (ed.), *The Foucault Reader*, 76-100. Harmondsworth: Penguin.
- Foucault, M. (1984b). 'Le Souci de la Vérité.' In Defert *et al* (éds.) (1994), *Dits et Écrits*, IV, 646-649.
- Foucault, M. (1984c). 'Qu' est-ce que les Lumières?'. In Defert *et al* (éds.) (1994), *Dits et Écrits*, IV, 562-578.
- Foucault, M. (1984d). 'What Is Enlightenment?' In Rabinow (ed.), *The Foucault Reader*, 32-50.
- Foucault, M. (1990). 'The Concern for Truth.' In L. D. Kritzman (ed.), *Politics, Philosophy, Culture: Interviews and Other Writings, 1977-1984*, 255-267. New York and London: Routledge.
- Foucault, M. (1991). *Discipline and Punish: The Birth of the Prison* [DP]. Trans. A. Sheridan. Harmondsworth: Penguin.
- Foucault, M. (1998). *The History of Sexuality*, Vol. 1: *The Will to Knowledge* [HSI (E)]. Trans. R. Hurley. London: Penguin.
- Foucault, M. (2001). *L' Herméneutique du Sujet: Cours au Collège de France (1981-1982)*. F. Gros (éd.), sous la direction de F. Ewald & A. Fontana. Paris: Gallimard/Le Seuil.
- Foucault, M. (2002a). *The Order of Things: An Archaeology of the Human Sciences*. London and New York: Routledge.
- Foucault, M. (2002b). 'Interview with Michel Foucault.' In J. D. Faubion (ed.), *Essential Works of Foucault, 1954-1984*, Vol. 3: *Power*. Trans. R. Hurley *et al*, 239-297. London: Penguin.

- Foucault, M. (2002c). 'So Is It Important to Think?' In Faubion (ed.), *Essential Works of Foucault*, Vol. 3, 454-458.
- Foucault, M. (2005). *The Hermeneutics of the Subject: Lectures at the Collège de France, 1981-82*. Ed. F. Gros, general editors: F. Ewald & A. Fontana. Trans. G. Burchell. New York: Palgrave Macmillan.
- Fraser, N. (1981). 'Foucault on Modern Power: Empirical Insights and Normative Confusions.' *Praxis International* 1(3): 272-287.
- Fulda, H. F. (1975). *Das Problem einer Einleitung in Hegels Wissenschaft der Logik*. 2e Auflage. Frankfurt a.M.: Vittorio Klostermann.
- Gadamer, H.-G. (1971). *Hegels Dialektik: Fünf Hermeneutische Studien*. Tübingen: Mohr.
- Habermas, J. (1985). *Der Philosophische Diskurs der Moderne: Zwölf Vorlesungen*. Frankfurt a.M.: Suhrkamp.
- Habermas, J. (1990). *The Philosophical Discourse of Modernity: Twelve Lectures*. Trans. F. Lawrence. Cambridge: Polity Press.
- Harris, H. S. (1997). 'Hegel's Correspondence Theory of Truth.' In G. K. Browning (ed.), *Hegel's Phenomenology of Spirit: A Reappraisal*, 11-22. Dordrecht: Kluwer.
- Hartmann, K. (1972), 'Hegel: A Non-metaphysical View.' In A. MacIntyre (ed.), *Hegel: A Collection of Critical Essays*, 101-124. Garden City, N.Y.: Anchor Books.
- Hegel, G. W. F. (1952). *Phänomenologie des Geistes [PhG]*. Ed. J. Hoffmeister. Hamburg: Felix Meiner.
- Hegel, G. W. F. (1977). *Phenomenology of Spirit [PhS]*. Trans. A. V. Miller. Oxford: Oxford University Press.
- Heinrichs, J. (1974). *Die Logik der <Phänomenologie des Geistes>*. Bonn: Bouvier.
- Houlgate, S. (1986 & 2004). *Hegel, Nietzsche and the Criticism of Metaphysics*. Cambridge: Cambridge University Press.
- Hutchings, K. (1999). *International Political Theory: Rethinking Ethics in a Global Era*. London: Sage.
- Hutchings, K. (2003). *Hegel and Feminist Philosophy*. Cambridge: Polity Press.
- Hyppolite, J. (1946). *Genèse et Structure de la Phénoménologie de Hegel*. Tomes 1 & 2. Aubier et Paris: Éditions Mouton.
- Kojève, A. (1947). *Introduction à la Lecture de Hegel*. Paris: Gallimard.
- Lauer, Q. (1976). *A Reading of Hegel's Phenomenology of Spirit*. New York: Fordham University Press.
- Maker, W. (1994). *Philosophy Without Foundations: Rethinking Hegel*. Albany: SUNY Press.
- May, T. (1993). *Between Genealogy and Epistemology: Psychology, Politics, and Knowledge in the Thought of Michel Foucault*. Pennsylvania: Pennsylvania State University Press.
- Owen, D. (1997). *Maturity and Modernity: Nietzsche, Weber, Foucault and the Ambivalence of Reason*. London and New York: Routledge.
- Owen, D. (2002). 'Criticism and Captivity: On Genealogy and Critical Theory.' *European Journal of Philosophy* 10 (2): 216-230.
- Pöggeler, O. (1973). *Hegels Idee einer Phänomenologie des Geistes*. Freiburg und München: Karl Alber.
- Prado, C. G. (2000). *Starting With Foucault: An Introduction to Genealogy*. 2nd ed. Boulder and Oxford: Westview Press.
- Rose, G. (1981). *Hegel Contra Sociology*. London: Athlone.

An Anthology of Philosophical Studies

- Rundle, B. (B.B.R.) (2005). 'Correspondence Theory of Truth.' In T. Honderich (ed.), *The Oxford Companion to Philosophy*, 2nd edition, 178-179. Oxford and New York: Oxford University Press.
- Sallis, J. (1995). *Delimitations: Phenomenology and the End of Metaphysics*. 2nd ed. Bloomington and Indianapolis: Indiana University Press.
- Taylor, C. (1993, c.1977). *Hegel*. Cambridge, New York and Melbourne: Cambridge University Press.
- Westphal, K. R. (1997). 'Harris, Hegel, and the Truth about Truth.' In Browning (ed.), *Hegel's Phenomenology of Spirit*, 23-29.
- Westphal, M. (1979). *History and Truth in Hegel's Phenomenology*. Atlantic Highlands, N.J.: Humanities Press International.

Part 2

Metaphysics, Logic and Philosophy of Language

CHAPTER TWELVE

The Individual Variability Problem*

Dimitria Electra Gatzia

Studies show that there are widespread intrasubjective and intersubjective color variations among normal perceivers (Kuehni 2004, Webster et al. 2000, and Malkoc et al. 2005). In this paper, I argue that such variations present difficulties, albeit not the same kind, for objectivism and subjectivism about color. In § 2, I discuss the difficulties individual color variations present for objectivists. I consider various moves they might make to accommodate such phenomena and argue that each is unsuccessful. In § 3, I discuss a version of color subjectivism, i.e., dispositionalism, and argue that it cannot provide an adequate solution to the individual variability problem. In § 4, I propose an alternative account of color, which is better suited to handle the individual variability problem.

Color Objectivism

It is common to think that colors are those properties we judge physical objects to have on the basis of the way they appear to us. Thus, the question that immediately confronts us is “What kind of properties are the colors?” or, to be more specific, “What kind of property is, say, greenness?” Objectivists¹ maintain that colors are perceiver-independent properties of physical objects.² Greenness, in this view, is a perceiver-independent property normal perceivers experience physical objects as having in typical viewing conditions. Since color experiences are said to be ‘about’ colors, the experience of something green is about greenness. In veridical experience, the color property one is aware of is instantiated in one’s environment. In non-veridical experience, on the other hand, the color property one is aware of is not instantiated in one’s environment. Illusions and hallucinations are often cited as paradigmatic

*A longer version of this paper was accepted for publication by *Philosophia*. The present version has been edited from that paper, and appears here with the permission of the editors of *Philosophia*.

¹I am defining ‘objectivism’ broadly here to be inclusive of any theories that attempt to identify phenomenal colors with perceiver-independent physical properties of objects. See Tye (2000) as well as Hilbert and Byrne (2003).

²Objectivists disagree about which physical properties should be identified with the colors. For example, Hilbert & Byrne (2003) argue that colors (for opaque objects) are *productances* since they *produce* (i.e., reflect, emit, or transmit) a specific proportion of incident light; McLaughlin (2003) argues that colors are physical *bases* of dispositions; and Armstrong (1968) argues that colors are wavelengths of light that objects emit or reflect.

examples of non-veridical experiences. In the case of illusions, the object does not have the color it appears to have to a perceiver. In the case of hallucinations, the object one is aware of is not instantiated in one's environment.

Individual color variations seem to present a serious difficulty for objectivism.¹ To see this let us assume that an object cannot have two color properties all over simultaneously.² An object O, for example, cannot *be* both unique blue and greenish-blue all over simultaneously. But, as it happens, O looks unique blue to one normal perceiver but greenish-blue to another. Since O is either unique blue or greenish-blue, but not both, at least one of the perceivers must be misperceiving its color. The question is which one? Since both perceivers have normal vision, there seems to be no non-arbitrary way of ruling out one of their color experiences as illusory. Unless there is a privileged class of normal perceivers, there is no non-arbitrary way of deciding which of the perceivers is misperceiving the color of the object.³ But there seems to be no such privileged class. (Nor is there a privileged class of viewing conditions). To be sure, the problem is not that it is impossible to specify normal perceivers (or viewing conditions). Rather, too many of them can be specified depending on our purposes.⁴ In what follows, I shall consider various responses available to objectivists and argue that none is successful in providing an adequate solution to the individual variability problem.

Deny that there are Intersubjective Variations

Objectivists might concede that intersubjective variations occur at the level of determinate colors (*e.g.* greenish-blue or greenish-yellow) but deny that they extend to determinable colors (*e.g.*, yellow or green).⁵ They might argue that although the object looks, say, unique blue to John and greenish-blue to Jane, both of their experiences are veridical because it looks blue to both. Since the object is blue and it looks blue to both, neither of them misrepresents its color. However, empirical evidence suggests that intersubjective color variations extend to determinable colors (Malkoc et al. 2005, Hardin 2004). An object that looks, say, blue to one perceiver could look green to another under the same viewing conditions. Assuming that the object can have only one color, only one of them is correctly representing the object's color; but which one?

¹Note that the individual variability problem can also be posed as a problem for representationalism (or intentionalism) about the phenomenal content of experience.

²As we will see in § 2.4, Kalderon denies this assumption.

³These perceivers have normal vision in the sense that they have passed every test that is used to detect color vision deficiencies. Hilbert and Byrne, however, argue that the objectivist can solve the problem by taking color vision to be normal in the teleological sense. I address their proposal in § 2.3.

⁴See Hardin (1988).

⁵See Tye (2006). Tye uses the terms 'fine-grained' and 'coarse-grained' instead of 'determinates' and 'determinables.' For replies to Tye's argument see Cohen *et al.* 2006 and 2007.

Since both perceivers have normal color vision, there seems to be no non-arbitrary way of ruling out one of their color experiences as illusory.

Countenance Unknowable Color Facts

Hilbert and Byrne (2003) argue that intersubjective color variations do not threaten objectivism since they merely establish that there are “unknowable color facts” (Hilbert & Byrne 2003). Admitting that there are unknowable facts is fairly unproblematic since it seems true that there are facts we cannot know. However, admitting that there are unknowable *color* facts seems to undermine objectivism. If colors are objective physical properties of objects visual experience represents, it seems reasonable to expect that we can know which perceivers correctly represent the colors of objects.

Objectivists might appeal to externalism about knowledge to argue that a perceiver *S* *can* know which color an object has (even though they might concede that we cannot tell which perceiver this is.)¹ In other words, even though objectivists might not be able to deny that there are unknowable color facts, they might nevertheless be able to use externalism about knowledge to restrict its domain. Externalism says that there is some sort of external condition that captures the proper relation between belief and truth. In particular, whether *S*'s belief about the color of an object amounts to knowledge depends on conditions external to *S*'s cognitive perspective. On Nozick's account, for example, *S* knows that *p* just in case the following conditions are met: *S* believes *p*, *p* is true, and if *p* were false, then *S* would not believe *p*. Beliefs that satisfy these conditions count as knowledge because they *track* the truth.²

Can externalism about knowledge, in this case Nozick's account, be used to block the conclusion that *S* cannot know that an object is unique green? Recall that normal perceivers locate unique green somewhere between 490 nm and 520 nm. Suppose that *S*, who is a normal perceiver, locates unique green at 515 nm and that this is the actual location of unique green. By hypothesis, the stimulus *x* is unique green. Would *S*, in such circumstances, believe that *x* is unique green? It depends on which counterfactual circumstances we consider, i.e., which worlds are the closest possible worlds. If we imagine only worlds where *S* continues to have reliable color perception, then presumably *S* would not believe that *x* is unique green if *x* were not unique green. Thus, on Nozick's account, *S* would know that *x* is unique green since each of the above conditions would be satisfied. However, if we consider worlds where *S* does not have reliable color perception, then *S* in such circumstances would continue to believe that *x* is unique green even if *x* were some other color.

¹I am indebted to an anonymous referee for this suggestion and to Christopher Buford for discussion.

²There is a fourth condition, namely, if *p* were the case, then *S* would believe that *p*. However, since it usually goes unmentioned, I will not include it here. See *Philosophical Explanations*. Cambridge (MA): Belknap Press, 1981.

Thus, on Nozick's account, S would not know that x is green. Which of these two scenarios is more plausible?

It seems to me that individual color variations generate cases that are similar to the fake barns case. In the fake barns case, S sees what looks exactly like a barn and thinks that he is seeing a barn. S, however, does not realize that there are many barn facades in the area. Had S been looking at one of them, he would have been deceived into believing that he was seeing a barn. S was lucky not to have done that. Most externalists would agree that in this case S's belief does not amount to knowledge. Similarly, in our example, S locates unique green at 515 nm and thinks that 515 nm is the location of unique green. What S does not realize is that unique green could have been located anywhere between 490 and 520 nm. Had S located unique green anywhere else in the array, he would have been deceived into believing that x is unique green. S was lucky not to have done that. If we deny that S's belief would amount to knowledge in the fake barns case, we should also deny that S's belief would amount to knowledge in the color variation case. If this is right, externalism about knowledge cannot be used to block the conclusion that S cannot know that an object is unique green. In general, externalism about knowledge cannot block the conclusion that normal perceivers (in normal circumstances) cannot know which colors objects have.¹

There is a related worry pertaining only to objectivists who are also committed to representationalism about color experience. According to Jackson (2003), Hilbert and Byrne's twin commitment leads to the implausible claim that we cannot know *what it is like* to have a color experience unless we know which physical properties are identical with which colors. If Jackson is right, such objectivists would have to countenance unknowable facts about the phenomenal content of color experience in addition to unknowable color facts.

Employ a Teleological Notion of 'normal' Perceiver

In a more recent paper, Hilbert and Byrne (2007) revisit the issue of intersubjective color variations. They argue that the problem of normal perceivers arises from the failure to distinguish between two distinct senses of 'normal', i.e., statistical and teleological. Since the statistical sense is the statistical average of many actual perceivers, they argue, it cannot be used to predict how an object's color would appear to a particular perceiver. They thus propose to replace it with the teleological sense. A perceiver's visual system functions Normally, in this sense, if it functions as designed to function by natural selection; a perceiver whose visual system functions Normally is said to

¹Even if there is some way out of this, there seems to be a further worry. Namely, appealing to externalism about knowledge might not be sufficient to establish that there is higher-order knowledge about color facts. In which case, not only we cannot know that S knows that x is red, but also S cannot know that S knows that x is red. Some argue that some sort of internalist account of knowledge is necessary for higher-order knowledge. If this is right, appeals to externalism will be insufficient to solve the problem.

be Normal. Once the statistical sense is replaced by the teleological it becomes apparent, Hilbert and Byrne claim, that philosophers mistakenly assume that there should be no color variations among Normal perceivers. Given the “astonishing complexity” of the system, it should not be surprising that it is often unreliable –that is, it often fails to accurately track color properties. Hilbert and Byrne liken color vision to speedometers and argue that individual color variations are as common and unproblematic as speed variations. Suppose that “John’s speedometer reads ‘30’ while Jane’s reads ‘32’” (p. 88). Such “minor misperceptions” about speed, they argue, are “perfectly common and unproblematic” since they result from the complexity of the system (p. 88).

It is understandable that widespread variations would result from the complexity of a system. However, it is not at all clear how the complexity of the system is supposed to explain that such variations are unproblematic. Hilbert and Byrne are silent as to why they think that the complexity of the system makes such variations unproblematic but the following seems to be a plausible explanation. In the every day context, a difference of 2 mph is insignificant and hence unproblematic because the standards of accuracy are low. For example, when drivers use their speedometers to monitor their speed, a difference of 2 mph is insignificant and hence unproblematic. In the scientific context, a difference of 2 mph seems to be unproblematic –even though the standards for accuracy are much higher –because there is a fact to the matter about which speedometer, if any, is precise. Similarly, Hilbert and Byrne argue that color variations arise from the astonishing complexity of the visual system:

It is hardly surprising that our visual apparatus is often unreliable when pushed to the limits of its resolution, and hardly surprising that it might calibrate slightly differently (p. 88).

Clearly such variations are widespread. But why are they unproblematic? Again, Hilbert and Byrne offer no explanation. Nevertheless, since visual systems are likened to speedometers, it is reasonable to think that the explanation would be along the same lines. In the every day context, it might be argued, color variations are too insignificant to either be noticeable or impede communication. So even though they are common, they are nevertheless unproblematic.

It is worth noting, however, that even substantial color differences often go unnoticed. As I mentioned earlier, the distribution of the spectral range of unique green between normal perceivers ranges from approximately 490 nm to 520 nm. As a result, a 503 nm stimulus will look yellowish to a perceiver who locates unique green at 490 nm but bluish to another who locates unique green at 515nm. Yet such differences often go unnoticed. Substantial speed differences, by comparison, become immediately apparent. How can this be? The reason seems to be twofold. Firstly, we tend to classify our color experiences under broad categories such as ‘green.’ Since color categories are far less precise than individual color experiences, even major differences tend to go unnoticed. Secondly, although there are roughly eight thousand color

terms in English, only approximately forty of them can be used reliably by trained perceivers. The consensus color vocabulary of ordinary English speakers is even more limited; it includes not many more than eleven basic color terms and their modifiers such as deep and light (Hardin 1988, Berlin & Kay 1969). Our limited color vocabulary thus compensates for the widespread intersubjective color variations. There are no such limitations, however, with respect to speed. Hence, the analogy fails.

What about the scientific context? Recall that minor speed differences, in the scientific context, were supposed to be unproblematic –even though the standards for accuracy are much higher –because there is a fact to the matter about which speedometer, if any, is precise. Similarly, color differences are supposed to be unproblematic –even though the standards for accuracy are much higher –because there is a fact to the matter about which visual system, if any, is accurate. Hilbert and Byrne argue that the teleological sense of ‘normal’ can be used to distinguish between Normal and abNormal perceivers. In the teleological sense, “Normal systems are those that function as “designed” by natural selection...and normal perceivers are those whose systems are Normal.”¹ Once the notion of ‘Normal visual systems’ is in place, objectivists can use it to distinguish between Normal and abNormal perceivers. They can then use this latter notion to distinguish between veridical and illusory color experiences. The color experiences had by Normal perceivers are veridical whereas the color experiences had by abNormal perceivers are illusory.

Hilbert and Byrne maintain that the colors can be identified with reflectance properties. The teleological notion of ‘normal’ is thus a natural offshoot of their account of color: a visual system functions as “designed” by nature just in case it accurately tracks reflectance properties.² Whether the teleological sense of ‘normal’ can be used to solve the individual variability problem thus depends on whether the function of the visual system is, in fact, to track color properties. Recent evidence from physiology, psychophysics, and neuroscience suggest that the function of color vision is not to track or detect colors but rather to recognize patterns such as objects and forms (Werner and Webster 2002, Gouras and Zrenner 1981).³ The fact that only a relatively small proportion of cells in our visual cortex have any wavelength selectivity, for example, suggests that “much of the neural machinery in the visual cortex is used for pattern recognition rather than color vision.”⁴ If this is right, it follows that the teleological sense of ‘normal’ (i.e., Normal) cannot be used to solve the individual variability problem.

¹Hilbert and Byrne (2007), footnote 2.

²Kathleen Akins (1996) coined this the ‘detectionist thesis’ of color vision.

³Even if the function of the visual system were to track color properties, objectivists would be compelled to adopt an anthropocentric view of color, which excludes non-human species, since color experiences across species lack unity. This is, in fact, what many do. For a defense of anthropocentric objectivism see Hilbert (1987). For arguments against color anthropocentrism see Matthen (1999). The account I defend in § 4, by comparison, is inclusive of other species.

⁴Peter Gouras’ ‘Color Vision.’ *Webvision: The Organization of the Retina and the Visual System*, John Morgan Eye Center, University of Utah. Available at <http://www.webvision.med.utah.edu/Color.html> [8 August 2008]. See also Gouras and Zrenner (1981).

Hilbert and Byrne could, of course, insist that the function of the visual system is to track colors, not patterns. But since this would be contrary to empirical evidence, they would have to argue either that the evidence is inadmissible or that it does not establish that the function of the visual system is not to track colors. In the absence of such arguments, the objection stands.

Accept Color Pluralism

Kalderon (2007) argues that intersubjective color variations present a problem for objectivists only under the presupposition that color properties are incompatible. We tend to think that an object cannot simultaneously be, say, unique green and yellowish-green because we falsely assume that unique green and yellowish-green are incompatible properties. Once this assumption is rejected, the individual variability problem vanishes.

Recall that the problem arises from the fact that an object can appear to have different colors to different normal perceivers. For example, it can appear unique green to John but yellowish-green to Jane. Since the same object cannot be both unique green and yellowish-green (all over) simultaneously, it follows that at least one of them is misrepresenting its color. The question is which one? According to Kalderon, neither of these perceivers is misrepresenting the object's color because it has *both* (p. 576-7). In general, color pluralism says that objects are multiply colored. The object appears both unique green and yellowish-green (all over) simultaneously because it *is* both unique green and yellowish-green (all over) simultaneously.¹ Since objects have multiple colors, there is no need to arbitrarily rule out some of the color experiences as illusory. Color pluralists thus settle the disagreement by maintaining that no (typical) color experience is illusory.

Although this account provides a solution to the individual variability problem, it does so at a cost. If an object can have as many colors as it appears to have to different normal perceivers, then no perceiver can be mistaken about its color. Color pluralism thus seems to be inconsistent with the possibility of error in color perception. It thus conflicts with the widely accepted claim that the possibility of error is a necessary aspect of the notion of representation. Kalderon acknowledges this difficulty and argues that his account is consistent, at least in principle, with the possibility of error in color perception (p. 582). For it can distinguish between veridical and non-veridical color experiences in some, although not in most, cases by reference to *local* sources of perceptual variation. Simultaneous contrast effects are sited as paradigmatic cases since they involve color shifts due to changes in the immediate surrounding.

Any achromatic surface, for example, can look black, white, or gray by increasing or decreasing the illumination of the area that immediately surrounds it (Hardin 1988). Colors whose appearances depend on their immediate surroundings are often called *contrast colors*. They include black,

¹Note that objectivists like Hilbert and Byrne would most likely not accept Kalderon's solution to the problem.

white, brown, navy blue, and olive. What distinguishes contrast colors from the rest is that their appearance depends solely on their immediate surroundings; their appearance changes when the illumination of the area that immediately surrounds them changes. For example, an object that looks brown under typical viewing conditions will look orange when viewed under the aperture mode of viewing, which can be created by eliminating the immediate surrounding.

Kalderon argues that when the source of variation is the result of simultaneous color contrast, *all* color experiences are illusory (or if one of them is correct it is only accidentally so). This purports to block the conclusion that color pluralism is not consistent with the possibility of error in color perception. On his account, when an object looks brown to a perceiver, her experience cannot be veridical (or, if it is, it is only accidentally so) because the object is not brown. However, such success comes at a cost. Color pluralism is made consistent with the possibility of error in color perception by ruling out as illusory a whole host of colors. In addition, Kalderon's argument assumes that the function of color vision is to detect or track *local* stimuli. This assumption, however, is false. As it has been shown, the function of color vision is to recognize patterns such as objects and forms. Rather than detecting *local* stimuli, the visual system recognizes patterns by analyzing the entire scene. The light that reaches the eye is thus a function of, among other things, the surface reflectance of the object, its surrounding surfaces, and various other variables such as illumination.¹

Color Subjectivism

The failure of objectivism to account for various phenomena has led many philosophers to embrace color subjectivism. Some color subjectivists such as Hardin (1988) argue that the fact that there is “no robust, perceiver-independent mapping from the physical to the phenomenal aspects of color” implies that colors are subjective properties (Thompson 1995, p. 124). Dispositionalists maintain that objects are colored but that colors are perceiver-dependent dispositions. Locke (1975), among others, seems to have embraced dispositionalism. He argued that colors “are in truth, nothing in the objects themselves” but “powers to produce various sensations in us by their primary qualities.”² In this view, a perceptible object is blue just in case it looks blue to normal perceivers in standard conditions.³

¹See Hardin (1988). Perhaps there are other examples that could be used to support Kalderon's argument. The point is that providing an adequate solution to the individual variability problem requires establishing that at least some of our typical color experiences are illusory. I have shown that the permissive nature of color pluralism is a hindrance to that end. For a more recent discussion on illusions, see also Kalderon's 'Color Illusion' forthcoming in *Noûs*.

²Locke 1975, Book II Ch. VIII. Maund (1995) argues that Locke's view is more complex since it requires that we distinguish between color as the property of physical bodies and color as a property of experience.

³See Harvey (2000). See also Peacocke (1984).

Dispositionalism has come under intense attack over the years.¹ The most common objection against it is that it seems to be circular: if the dispositionalist were to analyze ‘looks blue’ as *producing a color experience that represents its object as blue*, he would be explaining *being blue* by reference to *looking blue* and *looking blue* by reference to *being blue*. This objection, however, need not be fatal since the dispositionalist could explain ‘looks blue’ in terms of a sensory quality or phenomenal color experience.² Another, perhaps more serious, objection is that (first-order) dispositions cannot be causes. Since colors are causally efficacious and dispositions are not, colors cannot be identified with dispositions.³ The objection most relevant to the purposes of this paper –in the sense that it relates to the individual variability problem –is that it requires that we specify normal perceivers and standard conditions of viewing.⁴ Since blue, in this view, is the disposition to appear blue to normal perceivers under standard conditions, various difficulties will arise in attempting to specify these notions.⁵

One such difficulty involves metameric objects, i.e., objects that differ in their reflectance properties but appear to match perceptually under some illumination but not under another (Hardin 1988). Suppose a pair of metameric objects look to be blue under one illumination C_1 . Suppose also that under a different illumination C_2 one of these objects appears to be blue while the other appears to be greenish-blue. According to dispositionalism, the same object is both blue and greenish-blue since it appears to be blue in C_1 and greenish-blue in C_2 . This, however, is at odds with the intuition that an object has one of the colors it appears to have, not as many as it appears to have. In a recent paper, Harvey (2000) argues that metameric objects do not threaten dispositionalism. His argument, if cogent, could also be applied to intersubjective colors variations.

Harvey argues that the dispositionalist can handle metameric objects by adopting the notions “present color” and “official color”:

Dispositionalism can handle this [i.e. metamers] quite easily, providing that the dispositionalist account contains the key elements it should, and this includes the distinction between what I call an object’s “present colour” and its “official colour.” (p. 142)

An object’s present color, on the one hand, is determined by “the *present perceived colours of normal or standard conditions* in the object’s present situation” (ibid). The present color of an object changes when the viewing

¹For a list of critics and a defense of dispositionalism see Harvey (2000). For arguments against dispositionalism see Boghossian and Velleman (1989), Hardin (1983), and Averill (1982).

²See, for example, Boghossian and Velleman (1989).

³Jackson (2007), for example, argues that we see, and think of, color as a property of things in the world to which we are responding. When something looks blue to us, ‘we are in a state that is putatively a causal response’ to the object’s redness’ (170). See also Jackson’s ‘The Primary Quality View of Colour’, *Philosophical Perspectives*, 10: 199-219, 1669.

⁴The term ‘normal’ here is meant to be inclusive of the statistical and the teleological sense.

⁵See Boghossian and Velleman (1989).

conditions change. An object's "official colour," on the other hand, "is a matter of non-arbitrary convention, since it involves *designating some type of situation* to be the crucial type" (ibid). If daylight has tacitly been designated as the crucial type of situation, then the object's official color is that which it appears to have in that situation:

The official colour of a letterbox is red because ordinary daylight has tacitly been designated the crucial type of situation for determining the object's official colour (ibid).

The official designation purports to explain why the letterbox is not green when viewed under an artificial lighting that alters its appearance, e.g., it makes it appear green.¹ The relation between *present* and *official* colors is that the former has epistemological priority over the latter.

According to Harvey, "the significance of the perceived colour derives from *their rough uniformity among standard viewers*, and that uniformity is equally significant whatever the viewing conditions."² Harvey acknowledges that there are individual color variations among normal perceivers, but (falsely) assumes that this variability is limited to determinate colors such as scarlet and it does not extend to determinable colors such as green and blue. He argues, for example, that an object O is presently green if "nearly all" normal perceivers experience the object as any one of the shades belonging to a single color category designated as "green." Although O looks unique green to one normal perceiver and yellowish-green to another, it is nonetheless green because both shades belong to a single color category. As we have seen, the assumption that intersubjective color variations are limited to determinate colors is false. Individual variations extend to determinable colors such as green (Malkoc et al. 2005, Hardin 2004). The notion 'present colors' cannot thus be used to solve the problem of individual variations since which present color an object has depends on how it appears to "nearly all" normal perceivers.

The notion "official color" is equally problematic. Harvey maintains that the official color "involves *designating some type of situation* to be the crucial type" (p. 142). If daylight has tacitly been designated the crucial type of situation, then the object's official color is that which it appears to have in that situation. The letterbox is red because it appears red in ordinary daylight. Since objects appear to have different colors depending on what time during daylight they are viewed, the notion of "official color" would have to be severely restricted to a particular time during daylight. Despite the fact that most such color variations go unnoticed (partly because our color memories are fleeting) they nevertheless occur throughout the day.³ The notion "official color" would thus have to be severely restricted to apply to a particular time during daylight.

¹Harvey is responding to an objection made by Boghossian and Velleman (1989: p. 103) who offer such an example and argue that dispositionalism does not provide the correct answer to the question 'What color is the object?'

²Harvey (2000: p. 142) emphasis added.

³See Hardin (1998).

An object's official color, understood in this way, however, would not be "a matter of non-arbitrary convention" (p. 142). Suppose that an object O appears unique blue at 8 am but greenish-blue at 3 pm. Saying that O is unique blue, but not greenish-blue is arbitrary, and hence not a matter of non-arbitrary convention.

The success of Harvey's solution depends on whether there are non-arbitrary specifications of normal perceivers and standard conditions.¹ However, as we have seen, references to normal perceivers and standard conditions of viewing are problematic. To be sure, the problem is not that we cannot specify normal perceivers, but that too many of them can be specified depending on our purposes. Similarly, the problem is not that we cannot specify standards conditions,² but that too many of them can be specified depending on our purposes. It follows that the individual variability problem is a threat to dispositionalism. Although the phrase *blue things look blue to normal perceivers in standard conditions* appears to be a truism, as Hardin cautions us, "one should not take for granted that it can safely be pressed into heavy semantic duty" for instance, to specify in phenomenal terms exactly what it is for a physical object to be [blue]" (Hardin 1983, p. 806). Dispositionalism cannot succeed without reference to such notions. However, such notions are problematic, and so is dispositionalism.³

Color Fictionalism to the Rescue

As we have seen, there is considerable variation in color perception. Yet, neither of the theories I have discussed thus far can provide a satisfactory solution to the individual variability problem. Is there a theory of color that can accommodate such variability? In what follows, I shall argue that prescriptive color fictionalism is up to the task.⁴

Prescriptive color fictionalism is a species of error theory consisting of a *conceptual* and an *ontological* claim (Mackie 1977, Maund 1995 and 2005).⁵ According to the conceptual claim, ordinary color discourse genuinely purports to describe objective physical properties that cause color experiences. According to the ontological claim, there are no such properties. Color fictionalism concedes that color discourse entails or embodies a theory that is

¹Boghossian and Velleman (1989) also tie talk about normal perceivers and conditions of viewing to general agreement about determinable colors.

²See Hardin (1988). Note that Hardin used 'normal' in the statistical sense, but the argument can also apply to the teleological sense.

³For a comprehensive discussion of the Locke-Descartes type dispositionalism see Maund (1995: pp. 5-23). For a recent rejection of dispositionalism see Kalderon (2008).

⁴I defend this account in Gatzia 2010.

⁵*Prescriptive* fictionalism is distinct from *descriptive* fictionalism, which says that the target discourse (in this case, color discourse) is *already* treated as a fiction. Descriptive fictionalists maintain that although ordinary people seem to be expressing propositions that commit them to the existence of color properties when they employ color discourse, in actuality, they engage in some kind of pretense. For example, when ordinary people utter statements ascribing colors to objects such as 'Lemons are yellow,' they express propositions of the form 'In F, P.'

false. But it recommends that we can preserve color discourse by replacing assertion with the act of make-believe. In other words, it recommends that instead of asserting that objects are colored, we should pretend that they are.¹ (This does not change the content of statements ascribing colors to objects. It is the force with which we utter such statements that changes.²)

Color fictionalism has three components: a *base discourse*, a *fiction*, and *bridge principles*, which connect the former to the latter. The *base discourse* is the unproblematic, literally interpreted, part of the theory. It is also the crucial component in specifying correctness conditions. The *fiction* contains the false theory of color we ordinarily accept, namely that objects are colored. It is linked to the *base discourse* via *bridge principles*, which allow us to go from information about the world to conclusions about the same subject matter, taking a detour through the fiction.³

The base discourse incorporates “high-level statistical constructs built out of correlations between color experiences and other phenomena” (Johnson and Wright 2006: 140).⁴ Standard statistical methods can combine scientific data to extract such constructs from “statistical correlations between various kinds of environments and the color experience of various species” (Ibid, 159). The statistical model incorporated in the base discourse can be used to determine which experiences are erroneous: color experiences that fall within the cluster representing the majority of the population are readily endorsed. They can thus be treated as correct. The outliers, i.e., color experiences that fall outside of the cluster, on the other hand, deviate from the generally endorsed color experiences. They can thus be treated as erroneous.⁵ This framework can allow us to accommodate not only normal cases of perceptual variations but also hard cases.⁶ As Johnston and Wright (2006) note:

¹As Joyce (2001) notes, the act of make-believe differs from self-deception. When one utters a proposition *p* as an act of make-believe, one knows that *p* is false but pretends that it is true.

²Geach argues that a ‘thought may have just the same content whether you assent to its truth or not; a proposition may occur in discourse now asserted, now unasserted, and yet be recognizably the same proposition.’ The difference is the *force* with which such propositions are uttered. See ‘Assertion’, *Philosophical Review*, 74(4), 1965.

³See also Nolan, D., Restall, G., & West, C. ‘Moral Fictionalism versus The Rest’, *Australasian Journal of Philosophy*, 83(3): 307-329, 2005.

⁴I initially proposed that the base discourse consist of the opponent-process theory, but this approach did not yield correctness conditions (see ‘Fictional Colors’ (forthcoming) and Gatzia 2007).

⁵As I have argued, this sort of variability presents difficulties for objectivism, and subjectivism.

⁶Johnson and Wright argue that colors can be associated with high-level statistical constructs. However, they regard their ‘theory as ‘scientific’ rather than ‘philosophical’ because the theory is not designed to give the metaphysical essence of colors, or to provide a conceptual analysis of color, or to accomplish many of the other tasks that have been assigned to traditional philosophical theories of color.’ They explain that ‘rather than telling us what colors are, the theory expresses what science tells us about colors [since] at present we are not entitled to identify colors with some particular physical property, or any other relatively basic type of property. The best we can do is associate colors with a certain set of statistical regularities’ (p. 159). My account, by contrast, is philosophical since it aims to preserve color discourse, partly

Normal cases of perceptual variation pose no deep problem... [while]...the “hard cases” of variation will be absorbed by the degree of vagueness [i.e., variation] inherent in the scientifically appropriate individuation of the [high-level statistical constructs which will be associated with the] colors. There is no need...to make unprincipled hard choices in such cases. Allowing the latent variables extracted from the scientific data to drive the individuation of the colors [i.e., the high-level statistical constructs] gives us a proper standard against which we can judge whether or not a particular color attribution is correct (e.g., in terms of outliers) and it enables us to accommodate cases in which subjects disagree about their color attributions. (162-3)

The normal cases can be dealt easily since they concern perceivers who fall within the cluster. The ‘hard cases’ concern the outliers. (Imagine four quadrants divided by two axes. Perceivers within the cluster would be located in the upper positive quadrant while perceivers who are outliers would be located in upper negative quadrant.) There are two ways to deal with the hard cases. The standard way is to account for them independently from the cluster and to treat them as deviations from the norm. This seems to be what Johnson and Wright are suggesting when they say that the hard cases will be “absorbed by the degree of vagueness [i.e., variability] inherent in the scientifically appropriate individuation” of the color experiences within the cluster. This approach helps avoid error variance inflation (which is generated by the distance between the members of the cluster and the outliers) by isolating the error variance from the latent constructs, thus allowing the latent constructs to be understood in their purest form. Another, less orthodox, approach is to incorporate outliers in the cluster without generating error variance inflation.¹ Either way, color fictionalism can specify correctness conditions. It can thus provide an adequate solution to the individual variability problem, one that is consistent with the scientific findings. It follows that it fares better than the rest.

Conclusion

It has been argued that intersubjective color variations present a problem for objectivist and subjectivist accounts of color. The fact that neither (types of) theories can provide an adequate solution to the individual variability problem, suggests that an alternative theory of color is needed. I argued that color fictionalism is a better alternative since it can use statistical models in its

because it is useful, in the absence of colors. It allows us to link the scientific theory of color perception, which is literally true, with the fiction via bridge principles.

¹See Karen Larwin’s ‘Latent Variable Scores From SEM Analysis.’ Paper presented at the American Evaluation Association Annual Conference, November 14, 2009, Orlando Florida.

base discourse to specify correctness conditions, thereby providing a solution to the individual variability problem.

References

- Armstrong, David M. (1968) *A Materialist Theory of Mind*. London: Routledge.
- Akins, Kathleen (1996) 'Of Sensory Systems and the 'Aboutness' of Mental States.' *The Journal of Philosophy* 93(7): 337-372.
- Averill, Edward W. (1982) 'The Primary-Secondary Quality Distinction.' *Philosophical Review* 91(3): 343-461.
- _____. (1992) 'The Relational Nature of Color.' *The Philosophical Review* 101: 551-588.
- Berlin, Brent and Kay, Paul. (1999) *Basic Color Terms: Their Universality and Evolution*. CSLI Publications.
- Bimler, D. L., Kirkland, John, Jameson, K. A. (2004) 'Quantifying Variations in Personal Color Spaces: Are There Sex Differences in Color Vision?' *Color Research and Application* 29(2): 128 – 134.
- Block, Ned. 1999. 'Sexism, Racism, Ageism and the Nature of Consciousness.' In Moran, Whiting and Sidelle (ed.), *The Philosophy of Sydney Shoemaker, Philosophical Topics* 26(1, 2).
- Boghossian, Paul A. and Velleman, J. David. (1989) 'Colour as a Secondary Quality.' *Mind* 98 (389): 81-103.
- Cohen, Jonathan. (2003) 'Color: A Functionalist Proposal.' *Philosophical Studies* 113(1): 1-42.
- Cohen, J., Hardin, C. L., McLaughlin, B. P. (2006) 'True Colors,' *Analysis* 66 (292), 335-340.
- _____. (2007) 'The Truth about 'The Truth about True Blue.' *Analysis* 67(294): 162-166.
- _____. (2007). 'A Relationalist's Guide to Errors About Colour Perception', *Noûs* 41(2): 335–353.
- Clark, Austin. (2000) *A theory of Sentience*. Oxford: Oxford University Press.
- Gatzia, Dimitria Electra. (2007) *Color Fictionalism: Color Discourse Without Colors*, Ph.D. diss., Syracuse University.
- _____. (2010) 'Colour Fictionalism', *Rivista Di Estetica*, 43(1).
- _____. 'Fictional Colors,' *Sorites*, 21 (forthcoming).
- Gouras, P. and Zrenner E. (1981) 'Color vision: a review from a neuro-physiological perspective.' *Progress in Sensory Physiology* 1: 139-179.
- Hardin, C. L. (2004) 'A Green Thought in a Green Shade.' *The Harvard Review of Philosophy* 12: 29-39.
- _____. (1983) 'Colors, Normal Observers, and Standard Conditions.' *Journal of Philosophy* 80: 806-13.
- _____. (1988) *Colors For Philosophers: Unweaving the Rainbow*. Indianapolis: Hackett.
- Harvey, J. (2000) 'Colour-Dispositionalism and Its Recent Critics.' *Philosophy and Phenomenological Research* 61(1): 137-155.
- Hilbert, David. (1987) *Color and Color Perception: An Anthropocentric Realism*. Stanford: Stanford University, CSLI.
- Hilbert, D. and Byrne, Alex. (2007) 'Truest Blue.' *Analysis* 67(1): 87-92.
- _____. (2003) 'Color Perception and Color Science.' *Behavioral and Brain Sciences* 26(1): 3-21.
- Hering, Edward. (1964) *Outlines of a Theory of the Light Sense*, Leo M. Hurvich and Dorothea Jameson(trans.). Cambridge (MA): Harvard University Press.
- Johnson, Kent and Wright, Wayne. (2006) 'Colors as Properties of the Special

- Sciences.' *Erkenntnis* 64: 139-168.
- Joyce, Richard. (2001) *The Myth of Morality*. Cambridge: Cambridge University Press.
- Jackson, Frank. (2007) 'Colour for Representationalists.' *Erkenntnis* 66: 169-185.
- _____. (2003) 'Color and Content.' *Behavioral and Brain Sciences* 26(1): 34-34.
- Kalderon, Mark, Eli. (2007) 'Color Pluralism.' *The Philosophical Review* 116(4): 563-601.
- _____. (2008) 'Metamerism, Constancy, and Knowing Which.' *Mind* 117: 935-971.
- Kuehni, Rolf, G. (2004) 'Variability in Unique Hue Selection: A Surprising Phenomenon.' *Color Research and Application* 29(2): 158-162.
- Locke, John. (1965/1975) *An Essay Concerning Human Understanding*. Oxford: Clarendon Press.
- Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin Books.
- Malkoc, G., Kay P., Webster M. A. (2005) 'Variations in normal color vision. IV. Binary hues and hue scaling.' *Journal of the Optical Society of America A*. 22(10): 2154–2168.
- Matthen, Mohan. (1999) 'The Disunity of Color.' *The Philosophical Review* 108(1): 47-84.
- Maud, Barry. (2005) 'The Illusory Theory of Colour: An Anti-Realist Theory.' *Dialectica* 60(3): 245-268.
- _____. (1995) *Colours: Their Nature and Representation*. Cambridge: Cambridge University Press.
- McGilvray, James, A. (1994) 'Constant Colors in the Head', *Synthese* 100: 197-239.
- McLaughlin, Brian P. (2003) 'The Place of Colour in Nature.' In Mausfeld, R. and Heyer(ed.), *Color Perception: Mind and the Physical World*, D. Oxford: Oxford University Press.
- Neitz J, Carroll J, and Neitz M. (2001) 'Color Vision: Almost Reason Enough for Having Eyes', *Optics and Photonic News*.
- Neitz, Maureen, and Neitz, Jay. (1998) 'Molecular Genetics and the Biological Basis of Color Vision,' in *Color Vision: Perspectives From Different Disciplines*, Berlin: Walter de Gruyter.
- Palmer, Stephen E. (1999) *Color Vision: Photons to Phenomenology*. Massachusetts: The MIT Press.
- Pardo, Pedro J., Pe´rez A. L., and Suero M. I. (2007) 'An Example of Sex-Linked Color Vision Differences.' *Color Research and Application* 32(6): 433 – 439.
- Peacocke, Christopher. (1984) 'Colour Concepts and Colour Experience', *Synthese* 58 (March): 365-382.
- Seager, William and Bourget, David. (2007) 'Representationalism about Consciousness.' In Velmans and Schneider(ed.), *The Blackwell Companion to Consciousness* Susan. Oxford: Blackwell Publishing Ltd.
- Thompson, Evan. (1995) *Colour Vision: A study in Cognitive Science and the Philosophy of Perception*. New York: Routledge.
- Tye, Michael. (2000) *Consciousness, Color, and Content*, Cambridge (MA): MIT Press.
- _____. (2006) 'The Puzzle of True Blue.' *Analysis* 66(3): 173-178.
- Van Fraassen, Bas C. (1980) *The Scientific Image*. Oxford: Clarendon Press.
- Webster M. A., Miyahara E, Malkoc G, and Raker V. E. (2000) 'Variations in normal color vision. II. Unique hues.' *Journal of the Optical Society of America A*. 17:

1545 – 1555.

Werner, John S. and Webster, Michael A. (2002) 'Color Vision is Form and Object Vision', *9th Congress of the International Colour Association: Proceedings of SPIE* 4421: 10-15.

Wright, Wayne. 2003. 'Projectivism Representationalism and Color.' *Philosophical Psychology* 16: 515-533.

CHAPTER THIRTEEN

How to Avoid Relativism: What Forms the “Bedrock” that turns our Spade?

Patricia Hanna

Not empiricism and yet realism in philosophy, that is the hardest thing. *Remarks on the Foundations of Mathematics*, 325.

On a standard reading, it's not “the hardest thing”, it's impossible. I argue against the dominant, “uncontroversial” reading of Wittgenstein as a Relativist on the grounds that languages are possible only because the practices which constitute them answer to the constraints of a mind-independent world. Though not created by our decisions, we can access the world only via these practices. This is realism without empiricism. Practices and their application in the world give us the possibility of objectivity and access to truth. In turn, these practices and their function in our lives are what provide us with the final resting place of justification: the bedrock on which our spades turn.

Wittgenstein as Anti-Realist

28. ...one can ostensibly define...a numeral ...“That is called ‘two’—pointing to two nuts—is perfectly exact.—But how can two be defined like that? The person one gives the definition to doesn't know what one wants to call “two”; he will suppose that “two” is the name given to this group of nuts! — He may suppose this; but perhaps he does not. ... an ostensive definition can be variously interpreted in every case.

29. Perhaps you say: two can only be defined in this way: “This number is called ‘two’ ”. For the word “number” here shews what place in language ... we assign to the word ...

Here Wittgenstein seems to reject Ontological (or Metaphysical) Realism, and with it the possibility of ontic fidelity and truth. Our practices, not the world, determine what a word means or whether a proposition is true. As practices change, the meanings of words and the truth value of propositions also change: there is no truth outside practices. Thus, there is no role for the world in fixing meaning or truth-value. Any appearance of objectivity in the assessment of propositions is an illusion. What we thought was determined by a mind-independent world turns out to rest upon the *mind-dependent* interpretations which our “language games” supply. Even concepts as

fundamental to our understanding of the world as precision and exactness are our inventions.

88 I tell someone “Stand roughly here”— may not this explanation work perfectly? And cannot every other one fail too? But isn’t it an inexact explanation?—Yes ... Only let us understand what “inexact” means ... it does not mean “unusable” ...

...
“Inexact” is ... a reproach, and “exact” is praise ... that is to say that what is inexact attains its goal less perfectly than what is more exact ... the point here is what we call “the goal” ... single ideal of exactness has been laid down; we do not know what we should be supposed to imagine under this head—unless you yourself lay down what is to be so called ...

Here again, the mind-independent world has no role. Exactness isn’t in the world, we invent it, we decide whether and how to apply it.

At this point, one may be tempted either to acquiesce in relativism or to pursue a “sceptical solution”. Each of these eliminates the problem of explaining how one might have realism without empiricism by rejecting realism. This, I argue, would be a mistake: Wittgenstein offers a way of providing realism without empiricism.

Wittgenstein as Realist

Immediately before his remarks on exactness, Wittgenstein writes “... The sign-post is in order—if, under normal circumstances, it fulfills its purpose” (Wittgenstein 1968, §88).

Understanding this remark depends on recognizing that understanding and conveying meaning do not arise in a vacuum: they arise in the context of human aims and purposes, i.e., our form of life, which, in turn, is embedded in our shared environment. Like a sign-post, if it is to be “in order”, a language must serve our purposes. These two—purposes and shared environment—are what allow realism without empiricism.

To see how this works, consider the following example.

Measuring

I want to convert a room into an office. Since I always misplace my tape, I use my feet to measure the room. Placing my feet heel-to-toe and counting each iteration, it’s 10 Hanna-feet X 12 Hanna-feet. My current office is 11 Hanna-feet X 13 Hanna-feet; it has one more door than the proposed office, so I decide to make the move. Since I want put down new oak flooring; I buy a tape, and measure the room again: it’s 9 1/2’ X 13’.

A friend asks me how big the room is. We've known each other since high-school, so I say it's 10 Hanna-feet X 12 Hanna-feet. She isn't satisfied. At first, I think she has a problem visualizing a Hanna-foot, so I tell her it's 9 1/2' X 13'; but she still isn't satisfied. I switch to a metric tape; no luck. I ask my husband to check my measurements. She's still unhappy: "What are its real dimensions?"

At this point, I know she's not asking so she can buy me a Kilm carpet, we're in a full-fledged metaphysical discussion. It's not that my measurements are inaccurate; it's not that they're imprecise: my friend wants the dimensions of the room given from the standpoint of a mind-independent system of measurement.

But, what could such standpoint amount to? Whenever we ask for or give dimensions, we must do so in terms of some modulus. We may prefer one modulus over others, but preference doesn't make it better than the others. Neither does precision or exactness. The Hanna-foot isn't precise; it varies from time to time, and circumstance to circumstance. It serves some purposes well precisely because it is imprecise—and, for the same reasons, it fails others.

None of this, however, entails that it's inaccurate. If I count correctly, a measurement in Hanna-feet is accurate; if I lose count, it's inaccurate. This is simply to say that Hanna-feet are subject to the same standards of correctness as feet, meters or any other unit of measurement.

When we pick a modulus of measurement, we pick the practice within which to take our measurements. Only now can we measure the room, and check that measurement for accuracy. The choice of modulus is decided by such considerations as the desired degree of precision and the need for widespread understanding. Choice of one modulus over another does not, however, imply *absolute* superiority, *absolute* exactness, or *absolute* correctness. Measurements can be accurate or inaccurate, or more or less precise, but none of this depends upon there being some measurement which gives the final—the *real*—answer.

This eliminates empiricism, but has it done so at the cost of realism? I argue it has not.

Realism without Empiricism

Dimensions require a context, which includes reasons for wanting them, as well as a means for getting and giving them: these means are constituted by the modulus. Since moduli are relative to practices, the resultant ontic units—Hanna-feet, inches, meters, &c.—are praxial, which is simply to say that they are the creatures of the practice: thus, the ontology is a praxial ontology.

It follows that there are no dimensions outside some practice of measurement. Does it also follow that there can be no truth about dimension? In order for this to follow, one would have to hold that words have new meanings each time we use them, and, thus, that there is no way of checking a

measurement or comparing today’s measurement with yesterday’s. But this is neither entailed nor presupposed by praxial ontology.

Meaning is tied to context of use, which includes the speaker’s purposes and the audience’s resources. These purposes and resources can, and often do, hold steady across practices—in the present case, this is to say, across different moduli of measurement.

If someone asks “How wide is the room? ”, there are many possible answers. Some are true and some are false. If, for example, the room is 10 Hanna-feet wide, then it is not 9 Hanna-feet wide. The statement “The room is 10 Hanna-feet wide” is true, while “The room is 9 Hanna-feet wide” is false. However, while there are many possible answers (both true and false), not every *response* given to a question counts as a possible answer. Answers must be capable of being true or false within the practice, and not all responses meet this requirement. Some responses are ruled out by the chosen practice. “Red”, for example, is an inappropriate response—it’s not an answer—to the question “How wide is the room? ”

This is the key to understanding how practices determine both meaning and truth. If a practice did not specify what lies outside it, i.e., what is neither true nor false, what is inappropriate, one might argue that the meanings the practice provides are bound to that practice and have no currency outside it. It is only by drawing boundaries around what does and what does not count as an *answer* that practices fix meaning in a way that allows for truth/falsity.

Now consider Wittgenstein’s remark at §82 *Philosophical Investigations* and how it relates to the gnomic remark at §242 “If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgments”.¹

82 ...The fact that one measurement is right automatically excludes all others ... It isn’t a proposition which I put against reality as a yardstick, it’s a system of propositions.
‘I haven’t got stomach-ache’ may be compared to ... ‘These applies cost nothing’. The point is that they don’t cost any money, not that they don’t cost any snow or any trouble. The zero is the zero point of one scale ...since I can’t be given any point on the yardstick without being given the yardstick, I can’t be given its zero point either. ‘I haven’t got a pain’ doesn’t refer to a condition in which there can be no talk of pain, on the contrary we’re talking about pain...— Otherwise my proposition would mean something

¹The rest of this passage is pertinent.

...This seems to abolish logic, but does not do so.—It is one thing to describe methods of measurement, and another to obtain and state results of measurement. But what we call “measuring” is partly determined by a certain constancy in results of measurement.

The implications of this have been discussed to a certain extent, but I shall have more to say about them in §5.2 where I will use some of Wittgenstein’s remarks in *On Certainty* to help illuminate them.

like: my present state has nothing to do with a painful one
...there's no connection between them. But I mean precisely that
there is a connection between my present state and a painful one.

Saying that the world is conceptualized by human practices means this: Practices limit the options proffered to the world. The Hanna-foot, for example, doesn't give the world the option of determining whether the room is red. "The room is n Hanna-feet long" "has nothing to do with" color. The same is true for any modulus of linear measurement: the practice of measurement makes options like "is red", "is hot", &c. irrelevant to the truth or falsity of "This room is n Hanna-feet long". The practice limits what can be substituted for n to the number of Hanna-feet paced off in one direction. The same holds for all other units of linear measurement.

The case of "red" is one extreme of this boundary drawing. But what about something less obvious? Suppose that you ask me for the dimensions of a room in meters and I respond in Imperial-feet. Is my response an answer? Strictly speaking, no. However, we may be tempted to say that it "will do", and that to insist that it isn't an answer is to be overly fastidious. A certain confusion arises because you can take my response given in Imperial-feet and convert it to meters—no doubt cursing me at every stage of the conversion. We have a mechanism which explicitly allows to move from one modulus to another in the case of meters and Imperial-feet, and this leads us to think that Imperial feet and meters are part of some single Ur-system of practices, a system which, perhaps, captures the true nature of the world. But, as the Hanna-foot shows, none of this is warranted.

Suppose, however, that I do not use Imperial-feet, but use Hanna-feet. Here the claim that this is an inappropriate response—i.e., that it's not an answer—seems reasonable. It seems reasonable because I have not only not answered your question, but I haven't even given you anything you can use to get to the desired answer. Hanna-feet do not "convert", not to meters, not to Imperial-feet, not to any "standard" unit of measurement, full-stop. The response in Imperial-feet is as obviously inappropriate as the response "red": it has nothing to do with how wide the room is in meters.

But this does not show that the practice of measuring in Hanna-feet lacks standards of truth and falsity. This practice, like those of measuring in meters and Imperial-feet, is capable of accuracy, and it specifies what's relevant to truth and falsity. Counting a response such as "Red" as an answer to the question "How long?" is ruled out by the practice. If I accurately pace off 12, then the statement "This room is not 10 Hanna-feet long" is true. This doesn't say that the length of the room has nothing to do with 10 Hanna-feet; it has everything to do with it: a correct answer to "How long?" will not be "10".

What is Truth?

One criticism of this proposal is that it hasn't preserved truth because it fails to show a connection between the world and our way of talking about it which preserves ontic fidelity; hence any claims to Epistemic Realism are misguided. In other words, what I have presented as Epistemic Realism is an illusion, and Praxial Realism is nothing more than a disguised version of relativism, all the more dangerous because it is disguised.

The critic might seek an explanation for this hidden commitment relativism in my reliance on Wittgenstein's writings. In *On Certainty*, for example, Wittgenstein writes “At the foundation of well-founded belief lies belief that is not founded” (253). This would seem to make our justifications and assessments for truth entirely internal to some ultimately arbitrary set of decisions on our parts.

Read in this way, Wittgenstein sounds much like Quine in passages such as the following, all taken from “Ontological Relativity” (Quine 1969).

...reference *is* nonsense except relative to a coordinate system. In this principle of relativity lies the resolution of our quandary (48).

In practice we end the regress of coordinate systems by something like pointing. And in practice we end the regress of background language ... by acquiescing in our mother tongue and taking its words at face value” (49).

What makes ontological questions meaningless when taken absolutely is not universality but circularity. A question of the form “What is an F?” can be answered only by recourse to a further term: “An F is a G.” The answer makes only relative sense; sense relative to the uncritical acceptance of “G” (53).

To show that this charge of hidden relativism, along with the attendant charge of relativism against Wittgenstein, is unfounded, I argue that truth doesn't rest on ontic fidelity, and remarks like those in *On Certainty* don't undermine my claim to have developed a *Wittgensteinian* realism. I argue that these remarks of Wittgenstein's advance the claim that Praxial Realism meets the goal of realism without empiricism, and that the appeal to bedrock is supported by the same reasons cited here to develop Praxial Realism.

Why Truth doesn't Stand on Ontic Fidelity

Ontic fidelity consists in representing reality in terms of the elements into which it naturally divides. An accurate account of the world follows these divisions, just as Plato's good butcher follows the “objective articulation” of the animal, and does not “attempt to hack off parts” willy-nilly (*Phaedrus*, 265e). Absent something like this, it is supposed, there can be no realism.

Praxial Realism upends this picture of truth. What is needed to establish truth (and falsity for that matter) are criteria of correctness, and these are given by the practice in which the scale (the modulus) acquires its meaning. Since we can use a modulus in multiple contexts, it provides a stable and reliable means for comparison. As Wittgenstein notes at *PI* 242, "...It is one thing to describe methods of measurement, and another to obtain and state results of measurement. But what we call 'measuring' is partly determined by a certain constancy in results of measurement". The basis of linear measurement lies in such criteria and stability, not in a world divided into fixed units which we strive to represent. Measuring in Hanna-feet proves realism without empiricism by showing that accuracy and precision do not rest on a mind-independent world, but rather rest upon the practice's specified means of application of the concepts. By using a modulus of measurement which is by its very nature imprecise and cannot be translated into any more precise modulus, the example of Hanna-feet allows us to separate considerations of accuracy from those of precision and exactness.

As indicated earlier, Imperial feet and microns are inter-translatable. Since microns are a more precise modulus of measurement, it's easy to suppose that if an accurate measurement in inches diverges from an accurate measurement in microns, the former must yield to the latter. Therefore, we suppose that microns give a better—truer—representation of the world. If our aim is to locate the form—i.e., the modulus—of measurement which represents the world *as it really is*, it might seem that the most precise measurement must surely give the true dimensions of objects.

This is mistaken. A Hanna-foot cannot be translated into inches or microns because it just is imprecise. But imprecision excludes neither usefulness nor correctness: saying a room is *n* Hanna-feet wide is either true or false. It is as correct, objective and accurate to say that the room is *n* Hanna-feet wide, as it is to say that it is *x* microns wide. Of course, this means that neither modulus gives us special insight into the real nature of the room; such a notion of *real* as mind-independent—i.e., Ontic Realism—has no purchase.

Bedrock and Practices

But, have I really defended a version of realism without empiricism? If there is no *final* answer, how could there be *truth*? If what I have said so far is correct, then I have avoided relativism without appealing to empiricism in the form of a mind-independent world which comes divided in conceptual units prior to any human interaction with it. But, if this is so, then what are we to make of Wittgenstein's appeals to bedrock, the end of justification, ungrounded action and unfounded belief. Is he indeed committed to something along the lines of acquiescence in our "mother tongue", or is there another way of reading these passages which preserves my claim that Praxial Realism Wittgensteinian and avoids relativism?

Let's review some worrisome passages from *On Certainty*.

110 ... As if giving grounds did not come to an end sometime. But the end is not an ungrounded presupposition: it is an ungrounded way of acting.

253 At the foundation of well-founded belief lies belief that is not founded.

To these we might add the following.

194 With the word "certain: we express complete conviction, the total absence of doubt, and thereby we seek to convince other people. That is *subjective* certainty.

But when is something objectively certain? When a mistake is not possible. But what kind of possibility is that? Mustn't mistake be *logically* excluded?

These three passages are enough to make one wonder whether Wittgenstein had concluded that there was no answer to the question of when something is objectively certain. On the one hand, it seems that logic alone can provide a form of certainty which is not subjective (though surely this is itself a relative certainty if Quine is to be believed). On the other, seems clear that logic cannot decide whether I turned off the bathwater before leaving home. But, in either case, hasn't Wittgenstein simply abandoned his search for realism *tout court*?

Perhaps, but I think not. A far better reading of the passages shows that the bedrock upon which we turn our spades and upon which all our justifications and our notions of empirical certainty rest lies in our actions. But it is not just any action, it's those actions which arise from and engage our environment: in short they are human actions and thus spontaneous, but they are not *acte gratuit*. In turn, these practices allow us to describe, access and investigate—measure, test, and so on—the "world". Thus at 110 when Wittgenstein says that giving grounds ends in "an ungrounded way of acting", he is not saying that it ends in a random action, rather, he is saying that the end is an action as contrasted with a proposition: to end the giving of grounds, the action must engage the world through a practice.

The first piece of evidence of this can be found in *On Certainty* itself.

204 Giving grounds, however, justifying the evidence, comes to an end;—but the end is not certain propositions' striking us immediately as true, i.e., it is not a kind of seeing on our part; it is our acting, which lies at the bottom of the language game.

Actions are the basis of justification, and by extension truth. But Wittgenstein is clear that the actions at issue aren't simply bodily movements, which may or may not be coordinated with one another: the actions he's concerned with are those which lie at the "*bottom of the language game*", i.e., at the bottom of a set of practices. The actions in question, in short, serve to

provide the ground for criteria and standards. The actions are engaged actions. In order to form a language-game, the actions must be capable of providing criteria and standards which, themselves, must be public and stable.

It's important that we see this aright: he is not saying that language games (practices in general) constitute our actions. Rather he is saying that it is *in our actions* that our language games find their bases.¹ Taken in the context of his remarks in the *Philosophical Investigations* about our shared form of life, it is an easy step to the conclusion that it is what we are and how we live our lives that gives rise to many points of commonality in our actions, and, hence, in our languages and conceptual frameworks. Further, insofar as this is the case, it follows that there are common entry points into our practices—including our language games. None of this rests on any form of Platonism or Kantianism, however naturalized it may be.²

We can sum this up as follows.

1. Some of our actions constitute language-games.
2. They can do so because they engage, and are engaged by, the world, and can, therefore, form the basis for criteria and standards: in other words, these actions are the ground for the conventions which form the practice.

It is now clear just how important role of action actually is. To illustrate this, look at the following passage in *On Certainty*.

324 Thus we should not call anybody reasonable who believed something in despite of scientific evidence.

What does this mean? Does it mean that being reasonable requires always recognizing how the world is, as the Ontic Realist would have it? Does it mean that we must take a “scientific method” oath and pledge our allegiance to scientific method? Does it mean that the unreasonable believer fails or refuses to recognize how the world is? None of these is required or entailed by the view defended here. What it means is that belief—expressed in language—must reflect and cohere with those actions upon which the language in which they are expressed is grounded. Thus, for me to say that dinosaurs lived on Earth 6,000 years ago, and “mean it”, would require me to reject all of the other empirical claims I make: I would be the person of *On Certainty* §231 of whom Wittgenstein says “...I should not understand, for *this* reason: I would not know what such a person would still allow to be counted as evidence and what not”.³

¹This is related to the claim in Hanna and Harrison 2004 that practices engage the world. Here we are looking at it from the other end: this engagement is what makes the practices possible.

²See Hanna and Harrison (2004) and Hanna and Harrison (forthcoming, 2010) for a full treatment of the failure of McDowell's attempt to naturalize Kant.

³Nothing Wittgenstein or I say entails that the reasonable person can't be mistaken or even deceive herself. It does mean that, at some point, the mistakes and/or self-deception will

In other words, holding to a belief “in despite of scientific evidence” requires adjusting my actions to this “belief”, and this would require me to deny my shared humanity and my shared environment. Indeed, it would require me to reject the very language game in which this “belief” is formulated.¹ And all of this is entirely impossible: I would “speak”—make sounds—but no one would understand me.

Our practices arise from our actions, and it is only within these practices that we can begin to make sense of terms like “true”, “false”, “certain”, “justified”, “evidence” and so on. Following the adherent to ontic fidelity, one might be tempted to think that the standard for determining whether or not something can be used as evidence is whether or not it is true. But this cannot be. In order to be able to appeal to “certain truth” we must first be able to explain how we decide whether or not a proposition is true or false, and this requires that we understand what does and what does not count as evidence. Thus “What we call “a mistake” plays a quite special part in our language games, and so too does what we regard as certain evidence” (*On Certainty*, 196), not because it reflects the world, but because it lies at the heart of the language game.

199 The reason why the use of the expression “true or false” has something misleading about it is that it is like saying “it tallies with the facts or it doesn’t”, and the very thing that is in question is what “tallying” is here.

200 Really “The proposition is either true or false” only means that it must be possible to decide for or against it. But this does not say what the ground for such a decision is like.

This is the role for our “decisions”. The language game—or rather the logic of the language game, which includes what we will count as grounds, is a creature of our practices, but our practices are creatures of our actions. As I have already shown (see §§3 and 4), these actions are not arbitrary, they are constrained by what and where we are, and as I have argued in §5.1, this is all that is needed for Epistemic Realism.²

collapse in on itself, and our “ad hoc” adjustments—the re-mappings which form the heart of Quine’s arguments for indeterminacy—will come up against the bedrock of our actions, and the carefully preserved theory will yield. But it yields to our engagement through practices, not the a world independent of human decisions.

¹To be more precise, I should say “in which” I make the noises or write the marks, “Dinosaurs lived on Earth 6,000 years ago”, as strictly speaking, nothing is affirmed by this.

²Of course, the sort of Ontic Realism embodied in McDowell’s approach, along with the Sellarsian framework within which it is located, isn’t going to fade away without a fight. A recent attempt to salvage it can be found in Kukla and Lance 2009. Here the authors offer an account based on pragmatics rather than semantics as traditionally understood to give a basis for McDowell’s approach to understanding language. Kukla and Lance address questions of rationality and objectivity in the context of an account of language, specifically questions of how we use language to convey meaning and establish truth, which attempts to show that the most fundamental forms of speech are the vocative, exemplified by “Yo!”-claims: “...second-

Conclusion

By separating Ontic Realism from Epistemic Realism, Praxial Realism meets Wittgenstein's demands for a realism without empiricism. Praxial Realism reconciles the fact that we can only access the world through categories we invent with the possibility of correctness in our descriptions; it gives us objectivity without appealing to a mind-independent structure of reality. I have not claimed that Praxial Realism is to be found in Wittgenstein's writings, but I have given evidence to show that it is fully consistent with them, and that it is, in the sense of taking its inspiration from these writings, a Wittgensteinian view. More importantly, it makes clear how action constitutes the bedrock upon which language, truth and objectivity ultimately rest.

References

- Hanna, Patricia 2009, "Review of "Yo! " and "Lo! ": *The Pragmatic Topology of the Space of Reasons*", *Notre Dame Philosophical Review*, 3 July.
- Hanna, Patricia and Harrison, Bernard 2004, *Word and World: Practice and the Foundations of Language*, Cambridge: Cambridge University Press.
- Hanna, Patricia and Harrison, Bernard forthcoming 2010, "The Limits of Relativism in Late Wittgenstein" (co-author, Bernard Harrison), to appear in *the Blackwell Companion to Relativism*, edited by Steven Hales, Oxford: Blackwell.
- Harrison, Bernard and Hanna, Patricia 2003, "Interpretation and Ontology: Two Queries for Krausz," in Ritivoi, Andreea Deciu, ed. *Interpretation and Its Objects: Studies in the Philosophy of Michael Krausz*, Clarence, NY: Rodopi Publishers, 93–108.
- Kripke, Saul 1982, *Wittgenstein on Rules and Private Language*, Cambridge, MA: Harvard University Press.
- Kukla, Rebecca and Lance, Mark 2009, "Yo!" and "Lo!": *The Pragmatic Topology of the Space of Reasons*, Cambridge, MA and London: Harvard University Press.
- Quine, W. V. O. 1969, *Ontological Relativity and Other Essays*, New York and London: Columbia University Press.
- Wittgenstein, Ludwig 1996, *Remarks on the Foundations of Mathematics*, Revised Edition, G. H. von Wright, R. Rhees and G. E. M. Anscombe, eds., Cambridge, MA: MIT Press.
- 1968, *Philosophical Investigations*, trans. G.E.M. Anscombe, New York: The Macmillan Company.
- 1969, *On Certainty*, ed., G.E.M. Anscombe and G.H. von Wright, trans. Denis Paul

personal speech acts that hail or call ..." (134) and observatives, exemplified by "Lo!"-claims, which express receptive encounters that capture "...not a preconceptual 'given' but a conceptually articulated experience" (53). Although on the surface this seems sympathetic to my argument that we need to look to our actions in the world, and their role in grounding practices, especially linguistic practices, in fact Kukla and Lance beg the question in connection with objectivity at precisely the most fundamental point, and the very point at which Wittgenstein offers an explanation: viz., at the point where rejecting the act results in one's being unable to engage in any act of affirmation or denial. See Hanna 2009 for a sketch of an argument against this in Hanna 2009 based on the reasoning presented here and in Hanna and Harrison 2004 and forthcoming 2010.

*How to Avoid Relativism:
What Forms the "Bedrock" that turns our Spade?*

and G.E.M. Anscombe, Oxford: Blackwell Publishing.

CHAPTER FOURTEEN

Two Dimensions of Logical Theory: Commitment and Truth

John Kearns

At present, a logical system, or logical theory, is understood to have three components:

- (i) A specialized or formal language, usually artificial,
- (ii) A semantic account for the formal language, which usually explains the truth conditions of sentences in the language,
- (iii) A deductive system which codifies a class of logically distinguished items of the logical language.

There are different ways in which people regard, or think of, the logical formal language. It might be an ideal, or logically perfect, language. Or it might be a simplified version, a model, of a natural language. Or it might be intended to bring to light some logically significant features of our ordinary (natural) language. However, I find it most helpful to think of sentences in the logical language as representing statements that we do, or might, make with sentences of our actual language. We don't use sentences of the logical language to *make* statements or perform other speech acts, we use them to *represent* statements. The semantic account is really for the statements that are represented by sentences of the logical language. The logical theory is a tool for investigating certain aspects of our ordinary language.

A simple theory of propositional logic investigates the language L , which contains a number of basic, or *atomic*, sentences, perhaps infinitely many, and *compound* sentences formed from them with these *connectives*: \sim , \vee , $\&$, \supset . We don't actually produce the atomic sentences, but simply talk *about* them. We use upper case italic letters to represent sentences of the logical language: A , B , C ,... If A and B are sentences of the language L , then so are the following: $\sim A$, $[A \vee B]$, $[A \& B]$, $[A \supset B]$.

Since we don't actually produce the basic sentences of the language L , we can't say which logical-language sentences represent which statements. But we know that each logical-language sentence represents a statement that is either true or false. To present the semantic account for statements represented by sentences of L , we employ *interpreting functions* which assign either truth or falsity to each atomic sentence. An interpreting function gives one of the ways that things can turn out with our basic statements. Given a match-up between logical-language sentences and ordinary statements, only one

interpreting function is actually correct, but we don't know which one this is. (We couldn't, since we haven't produced the actual atomic sentences, or matched them to ordinary statements.) So we consider all of the interpreting functions.

If we have an interpreting function f which gives a truth-value to each atomic sentence of L , this function determines a (*truth-conditional*) valuation of the whole language L . This is a *truth-table* valuation, because familiar truth-tables are used to determine the values of compound sentences (and the statements they represent), once the values of component sentences (statements) are determined.

Given our interpreting functions and the valuations which they determine, we can define concepts of *logical truth*, *implication*, and *inconsistency* for the language L .

A sentence A of L is *logically true* if, and only if (from now on: iff) A is true for the valuation determined by every interpreting function of L .

A number of sentences A_1, \dots, A_n of L *imply* a sentence B of L iff there is no valuation determined by an interpreting function of L for which A_1, \dots, A_n are true but B is false.

A set X of sentences of the language L *imply* a sentence B of L iff there is no valuation determined by an interpreting function of L for which all sentences in X are true but B is false.

A set X of sentences of L is (*semantically*) *consistent* iff there is a valuation determined by an interpreting function of L for which all sentences in X are true.

A deductive system for L might be concerned with logical truth. This system would provide techniques for establishing that a sentence is logically true. If the system were entirely satisfactory, it would enable us to establish, for every sentence that is logically true, *that* it is logically true. A deductive system might also be concerned with *premiss-conclusion sequences*. So it would provide deductive techniques for establishing that the premisses of certain sequences imply the conclusions of those sequences.

The Logic of Speech Acts

I have been describing a standard system of logic. In 1985, John Searle and Daniel Vanderveken introduced the subject of illocutionary logic in their book *Foundations of Illocutionary Logic*. They regard this subject, which investigates illocutionary acts and their relations, as a kind of appendix to standard logic. But, in fact, the field that they introduced is broader than standard logic, and includes standard logic as a proper part. My own >take= on illocutionary logic, and the way that it should be investigated and developed, is quite different from theirs, but it is their idea that I am building on. (Some papers where I develop systems of illocutionary logic are listed in the references.)

My work in illocutionary logic is based on the understanding that speech acts, or language acts, and the skills and dispositions to perform language acts possessed by a community of language users, constitute the fundamental linguistic reality. Someone who uses an expression meaningfully performs a language act. That someone might speak out loud, or write something, or simply think in words. A person who reads or who listens with understanding is also performing language acts, but we usually focus on language acts from the perspective of the person who produces the expressions that are used.

I use the expressions >speech act= and >language act= interchangeably, but >language act= is often the more appropriate expression, because many speech acts are not actually spoken aloud. Speech acts are the primary bearers of semantic features such as meaning, truth, and truth-conditions. Spoken or written expressions are the bearers of syntactic features, and can themselves be regarded as syntactic objects. For logical investigations, *sentential acts*, or acts performed by using a single sentence, are especially important. Those sentential acts that are true or false are even more important. I understand a *statement* to be a true or false sentential act; this is a *stipulated* meaning for this word, for statements are more often understood to be something like assertions.

Sentential acts are typically performed with a certain *illocutionary force* such as the force of a promise, of an assertion, a question, or a request. Statements themselves can be performed with a variety of illocutionary forces. A single statement can be *asserted* or *denied*, it can be *supposed true* or *supposed false*. I will take an *assertion* to be an act of making a statement (which is Aperforming@ the statement) and coming to *accept* it as being or representing what is the case, or an act of making a statement and reaffirming one=s continued acceptance of the statement. This is also a stipulated meaning for >assertion,= since my assertions don=t require an audience, and are always sincere. A *denial* is an act of rejecting a statement for failing to represent what is the case, or an act indicating one=s continuing rejection of the statement.

A *supposition* is an act of temporarily, or provisionally, accepting a statement, or rejecting a statement, to determine what follows from that supposition when it is combined with other illocutionary acts that we perform. If a statement is supposed true, or false, and some conclusion is deduced from it, that conclusion also has the status of a *supposition* and I shall call it a supposition.

Arguments as I shall understand them are language acts in which a person begins with premisses which are illocutionary acts, and reasons to a conclusion which is also an illocutionary act. Premisses and conclusions are not simply statements, they are acts of accepting statements, or of rejecting them, etc.

A system of illocutionary logic is obtained from a standard system by making three changes:

- (i) Illocutionary-force indicating expressions, *illocutionary operators*, are added to the formal language.

- (ii) The semantic account of truth-conditions is supplemented with an account of semantic features of illocutionary acts.
- (iii) The deductive system is modified to accommodate illocutionary operators. This makes it possible for the system to produce, and to explore, genuine speech-act arguments. Standard deductive systems are limited to tracing truth-conditional connections linking statements, by means of what I call *deductive derivations*, which are not arguments. The components of genuine arguments are illocutionary acts, and not simply statements.

Plain and Completed Sentences

An illocutionary logical system, or theory, contains a standard logical theory as a proper part. The standard theory is the *ontological*, or *ontic*, component of the illocutionary theory, for the standard theory considers statements for representing what there is, and determines what are the truth conditions of these statements. In the standard system described earlier, the atomic and compound sentences of the language L are *plain sentences* of the language L . The language L contains the following *illocutionary operators*:

$$\begin{array}{ll} \vdash & \text{B the sign of assertion} & \dashv & \text{B the sign of denial} \\ \sqsubset & \text{B the sign of positive supposition} & \neg & \text{B the sign of negative} \\ & & & \text{supposition} \end{array}$$

A *completed sentence* is the result of prefixing a plain sentence with an illocutionary operator. Plain sentences represent statements apart from illocutionary force, while completed sentences represent illocutionary acts.

Statements are true or false. The ontic component of an illocutionary logical theory deals with statements, their truth conditions, and with semantic properties defined in terms of truth and truth conditions. But illocutionary acts are not true or false, except in a derivative sense. So what are the semantic features of our four kinds of illocutionary act that can appropriately be studied, and investigated, in illocutionary logic? When someone comes to accept statements and to reject statements, these acts or their continuations *commit* her, they *rationally* commit her, to accept or reject further statements. If, for example, I assert that either Bill or Carolyn has recently bought a new car, but deny that Carolyn bought a car, then I am rationally committed to accept that Bill is the one with a new car. Supposing statements to be true, or false, commits a person to make further suppositions.

It isn't simply language acts that are linked by rational commitment. If I decide to do, or not to do, something, my decision commits me to do it, or not to do it. Many intentional acts generate commitments to perform, or not perform, other acts. Rational commitment can be conditional or unconditional. I may be rationally committed to close the upstairs windows if it rains while I am at home. This commitment does not take effect until it rains while I am

at home. But my decision to give a take-home final to my logic class next semester produces an unconditional commitment. A rational commitment is not, as such, a moral commitment. If I decide to buy a bottle of wine on the way home from work, but then forget to do it, I have done nothing wrong in failing to honor my commitment. And commitments can be abandoned as well as taken on.

When I accept some statements and reject others, this commits me to accept and to reject further statements. But the commitment is conditional. I am committed to do this if I have some interest in the matter, and give it some thought. I have accepted, and continue to accept, many statements, and similarly for denials. These commit me to accept, or to reject, indefinitely many more statements, but I have no interest in most of these consequences. However, it would be irrational to accept a statement $\supset[A \vee B]$ and reject B , but refuse to accept A .

Coming to accept or to reject a statement is performing an action. Once I have accepted a statement, I continue to accept it until I either forget or change my mind. Continuing to accept a statement is not an action, it is being in a certain state. I can be committed both to come to accept a statement, and to continue to accept a statement. And both coming to accept (or to reject, etc.) and continuing to accept a statement will commit me to accept or to reject further statements.

At the ontic level of an illocutionary logical theory, or in a standard logical theory, we explore semantic features like implication and inconsistency, which are defined in terms of truth conditions. There are counterpart features at the epistemic level of an illocutionary theory. If the person who performs the illocutionary acts represented by completed sentences A_1, \dots, A_n must be committed either to come to be or to continue to be in the state indicated by completed sentence B , then A_1, \dots, A_n *logically require* B , and the illocutionary acts that are represented are also linked by logical requiring. And if A is a plain sentence of L , then the following pairs of completed sentences (and the illocutionary acts these represent) are *logically incoherent*:

$$\vdash A, \dashv A \qquad \perp A, \neg A$$

Illocutionary acts which logically require incoherent acts are themselves incoherent.

Although logical requiring is an epistemic-level counterpart to implication, there are important differences between the concepts. There is no distinction between simple implication and complex implication, so that complex implication is built up from simple implication. But commitment is different. We *do* distinguish *immediate* commitment from *mediate* commitment. The immediate commitments of an illocutionary act or its continuation are evident to the person who performs the act. Mediate (remote) commitment is discovered (constituted) by chaining together immediate commitments. The natural way to determine the remote commitments of

illocutionary acts is by inference/argument from initial illocutionary acts to their consequences.

Some illocutionary acts, when they are first performed, create, or generate, commitments. For example, coming to accept statement A generates the commitment to perform this act:

$\vdash[A \vee B]$ (for any statement B). Other illocutionary acts *reveal*, rather than generate, commitments. Someone who uses a name to refer to an individual reveals that she is committed to assert that the referent exists. Whether performing an illocutionary act C generates, or reveals, a commitment to perform act D , performing C *logically requires* performing D .

Some Epistemic Semantic Concepts

Plain sentences, or the statements which these represent, are true or false. In studying statements and their semantic features, we employ *interpreting functions* which assign T and F to atomic plain sentences of the logical language. To give a semantic account of illocutionary acts represented by completed sentences of the logical language, we must notice a crucial difference between statements and illocutionary acts. In studying statements and their features, we pay little attention to who it is that makes the statements. For essentially the same statement can be made by different people, and the semantic properties of the statement are independent of the person who makes it. In dealing with illocutionary acts, we must consider the persons performing these acts. Smith=s assertion commits him to make further assertions and denials, but makes no claims on Brown. And Smith is not committed by the assertions and denials made by Brown.

To develop a semantic account for completed sentences of the language L , we must adopt the perspective of a particular person, and develop the account for, or with respect to, that person. I will consider an idealized person, a woman whom I call the *designated subject*. But the account which fits her and her illocutionary acts will also fit each of us and our acts, if we consider it from our own perspectives. We consider a given time, when the designated subject has accepted, and continues to accept, a number of statements, and has rejected, and continues to reject a number of other statements. These assertions and denials, together with her general linguistic competence, commit her at that time to accept and to reject a number of further statements. I use the plus sign $>+=$ for the Avalue@ of those assertions and denials which she has either performed or is committed to perform at the given time. In developing a logical semantic account, I employ functions which I call *commitment valuations*. These assign the value + to some completed sentences $>\vdash A=$ and $>\vdash B=$ of the logical language.

I treat the designated subject as an idealized language user in order to determine appropriate criteria for evaluating arguments made with completed sentences of L (to determine appropriate criteria for evaluating arguments made with illocutionary acts). We presume that the statements she has accepted are

true, and that she will continue to accept them, and the statements she has rejected are false. If an assertion $\>\vdash A=$ or denial $\>\dashv B=$ has value + at one time, it will also have that value at later times. This enables us to determine which arguments lead to conclusions where true statements are asserted or false statements denied, on the presumption that the arguments begin with assertions of true statements and denials of false statements. Commitment valuations do not assign the value + to suppositions, for suppositions come and go too quickly. We often begin an argument by supposing that a statement is true, and *discharge* that supposition before we finish the argument. The commitments of suppositions are explored by employing (truth-conditional) interpreting functions.

To formally characterize the epistemic semantic relations, we need some definitions. The first one involves a relation between commitment valuations and interpreting functions. A commitment valuation V is *based on* an interpreting function f iff all the assertions $\>\vdash A=$ assigned + by V are of statements A which are true for f , and all the denials $\>\dashv B=$ which are assigned + by V are of statements B which are false for f . A commitment valuation is *coherent* iff it is based on an interpreting function.

Given a coherent commitment valuation V_0 , we can imagine that this assigns + to those assertions and denials which the designated subject has actually performed by the time in question, which she remains committed to perform. This commitment valuation records her *explicit* beliefs and disbeliefs at the time in question. Another commitment valuation, V , assigns + to (exactly) those assertions and denials that are logically required by her explicit beliefs and disbeliefs. V is the *completion* of V_0 . A more careful definition follows.

Let V_0 be a coherent commitment valuation. The *completion of V_0* is the commitment valuation V such that (1) $V(\vdash A) = +$ iff A has value T for every valuation determined by an interpreting function on which V_0 is based, and (2) $V(\dashv B) = +$ iff B has value F for every valuation determined by an interpreting function on which V_0 is based.

Let V_0 be a coherent commitment valuation for L , and V be its completion. Let A be a completed sentence of L which is an assertion or denial. Then V_0 *satisfies* A iff $V(A) = +$.

Let f be an interpreting function for L , and let A be a plain sentence of L . Then (1) f *satisfies* $\vdash A$ iff $f(A) = T$, and (2) f *satisfies* $\dashv A$ iff $f(A) = F$.

Let f be an interpreting function for L , and let V be a commitment valuation based on f . Then $\langle f, V \rangle$ is a *coherent pair*.

Let $\langle f, V \rangle$ be a coherent pair, and A be a completed sentence of L . Then $\langle f, V \rangle$ *satisfies* A iff (1) A is an assertion or denial, and V satisfies A , or (2) A is a supposition, and f satisfies A .

Let A_1, \dots, A_n, B be completed sentences of L . Then A_1, \dots, A_n *logically require* B iff (1) B is an assertion or denial, and every coherent commitment valuation that satisfies the assertions and denials among A_1, \dots, A_n also satisfies B , or (2) B is a supposition, and every coherent pair which satisfies each of A_1, \dots, A_n also satisfies B .

Let X be a set of completed sentences of L , and B be a completed sentence of L . Then X *logically requires* B iff either (1) B is an assertion or denial, and every coherent commitment valuation which satisfies the assertions and denials in X also satisfies B , or (2) B is a supposition, and every coherent pair which satisfies every sentence in X also satisfies B .

Let X be a set of completed sentences of L . X is *logically coherent* iff there is a coherent pair which satisfies every member of X .

Our formal definitions of logical requiring and logical coherence adequately characterize these concepts with respect to our logical formal language, and provide targets to be captured or realized by our epistemic-level deductive systems.

Natural Deductions

By using commitment valuations to explore features of completed sentences, we have adapted techniques used to explore semantics in standard systems of logic to deal with illocutionary acts and illocutionary force. The fact that we can use similar methods to deal with both the ontic and the epistemic levels of a system of illocutionary logic is a good enough reason for regarding illocutionary force as a semantic feature of language (and language acts) rather than a pragmatic one. Language users *mean* both the content and the force of their illocutionary acts, and sentences are designed for performing acts with one or another force.

The natural way to explore commitment in a linguistic setting is by making inferences and constructing arguments. *Deductive reasoning* is reasoning that traces the *immediate* commitments of assertions, denials, and suppositions. Standard systems of logic, even when they employ what are called natural deduction systems, fail to recognize and accommodate illocutionary force and illocutionary acts. So-called systems of natural deduction introduce and discharge hypotheses, and in doing this, they mimic the practice of making, and discharging, suppositions. But they don't mark illocutionary force. Instead they focus entirely on statements and their truth-conditional features. In contrast, illocutionary logical theories provide the resources for making, and studying, genuine deductive arguments.

For example, in a standard system of logic, the rule/principle & *Introduction* can be illustrated as shown here:

$$\begin{array}{c} A \quad B \\ \hline [A \ \& \ B] \end{array}$$

But in a system of illocutionary logic, those sentences need to be prefixed with illocutionary operators. There are lots of ways to do this, but many of these result in A moves that are incorrect. Some of these are shown:

$$\frac{\perp_A \quad \perp_B}{\vdash[A \& B]} \qquad \frac{\perp_A \quad \vdash B}{\vdash[A \& B]} \qquad \frac{\vdash A \quad \perp_B}{\vdash[A \& B]}$$

It isn't just truth conditions that matter for a correct deductive argument. Illocutionary force must also be considered. A supposed premiss won't support an asserted conclusion. Here are correct instances of *& Introduction*:

$$\frac{\vdash A \quad \vdash B}{\vdash[A \& B]} \qquad \frac{\perp_A \quad \perp_B}{\perp[A \& B]} \qquad \frac{\perp_A \quad \vdash B}{\perp[A \& B]} \qquad \frac{\vdash A \quad \perp_B}{\perp[A \& B]}$$

For *& Introduction*, the forces must be positive, and the force of the conclusion must not exceed the force of the premisses.

If we represent simple arguments as indicated above, we can combine simple arguments to form complex arguments which have the shapes of tree structures:

$$\frac{\vdash[A \& [B \& C]]}{\vdash[B \& C]} \qquad \frac{\vdash[B \& C]}{\vdash B} \qquad \frac{\vdash[A \& [B \& C]]}{\vdash A}$$

$$\vdash[B \& A]$$

This argument establishes that the assertion $\vdash[A \& [B \& C]]$ logically requires this one:

$\vdash[B \& A]$. An illocutionary logical theory can make clear how both truth conditions and illocutionary force contribute to the correctness of inferences and arguments.

A failure to recognize illocutionary force is in part responsible for the *Amystery* behind indicative conditional sentences, and rules governing their use. It is commonly observed that the material conditional of modern logic isn't quite a good fit for ordinary conditionals, but there is considerable controversy about the exact character of the mismatch, and about what it would take to provide a good fit. The real source of this problem (according to me) is that ordinary conditional sentences are typically used to make *conditional assertions*, which are a kind of illocutionary act, rather than a kind of statement. Someone who uses a sentence $\text{>If } A, \text{ then } B$ to make an assertion is typically *asserting } B, \text{ on the condition (of) } A*. Since assertions aren't true or false, and conditional assertions are not assertions of conditional statements, it is no wonder that material conditional statements provide a less than perfect account of conditional assertions. In coming to accept one statement on the condition of another, what a speaker does is to establish an immediate commitment, or inferential connection, from asserting or supposing true the

antecedent to asserting or supposing true the consequent. An illocutionary logical theory of conditional assertions is an empirical theory of our actual linguistic practice.

Going Beyond Truth Conditions

There are many parallels between implication and logical requiring. If statements (or plain sentences) A_1, \dots, A_n imply B , then all of the following:

$$\begin{array}{l} \vdash_{A_1, \dots, A_n} \text{ logically require } \vdash B \\ \perp_{A_1, \dots, A_n} \text{ logically require } \perp B \end{array}$$

A mixture of assertions and positive suppositions of A_1, \dots, A_n logically require $\perp B$

So if statement A implies B , then $\vdash A$ logically requires $\vdash B$. But we can't say this: If $\vdash A$ logically requires $\vdash B$, then A implies B .

To see why not, let $\mathbf{>B=}$ be a first-person belief operator. So, for the designated subject, a sentence $\mathbf{>BA=}$ is true iff she has considered A and accepts or has accepted (and continues to accept) A . Then it is evident that the premiss act of the following inference principle logically requires the conclusion act:

$$\frac{\vdash A}{\vdash \mathbf{BA}}$$

and that this premiss act requires its conclusion:: $\frac{\vdash \mathbf{BA}}{\vdash A}$

but in neither case does the premiss statement imply the conclusion statement. Reflecting on the first of these principles makes clear what is wrong with the puzzling statement that figures in Moore's Paradox. That statement has this form: $[A \ \& \ \sim \mathbf{BA}]$. While such a statement may very well be true, the designated subject can't coherently accept it. For this completed sentence:

$$\vdash [A \ \& \ \sim \mathbf{BA}]$$

is logically incoherent. The assertion logically requires both $\mathbf{>BA=}$ and $\mathbf{>\sim BA=}$

That there are true statements which a given person can't coherently accept may be disappointing, but it is an inescapable limitation on belief and knowledge. Moore's Paradox depends on this fact, as does the paradox of the surprise execution (or examination). The limitation on knowledge (as opposed to belief) is sometimes attributed to Fitch. The limitation on both knowledge and belief is a logical limitation, but the logic involved is illocutionary logic

rather than standard logic. And the limitation isn't particularly damaging. I can't coherently both believe and disbelieve a single statement, but what I don't believe now I can believe later. And I can't now know which are the true statements that I don't know. But when I come to know one of them, I can also know that I didn't know it previously.

Expanding our understanding of what counts as semantic to include illocutionary force and the meanings of illocutionary force-indicating devices narrows our understanding of the pragmatic. My own suggestion is that pragmatics be understood as the study of how language is used to communicate meaning. What a speaker means by what she says, whether this is conventional meaning or a matter of conversational implicature, belongs to the semantic. How the addressee determines the reference of a given pronoun, the force of a given utterance, or the intended meaning of an ambiguous expression is in the domain of the pragmatic.

References

- Kearns, John T. (1997). Propositional Logic of Supposition and Assertion, @ *Notre Dame Journal of Formal Logic* 38, 325-349.
- _____. (2000). An Illocutionary Logical Explanation of the Surprise Execution, @ *History and Philosophy of Logic* 20, 195-214.
- _____. (2005). The Logical Difference between Knowledge and Justified Belief, @ *The Logica Yearbook 2005*, M. Bilkova and O Tomala eds., Filosofia, Prague, 101-112.
- _____. (2006). Conditional Assertion, Denial, and Supposition as Illocutionary Acts, @ *Linguistics and Philosophy* 29, 455-485.
- _____. (2007). An Illocutionary Logical Explanation of the Liar Paradox, @ *History and Philosophy of Logic* 28, 31-66.
- Vanderveken, Daniel and John Searle. (1985). *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

CHAPTER FIFTEEN

In what Sense is Tarski's Semantic Conception of Truth Semantic?

Ladislav Koren and Jana Pavlikova

In his path-breaking work, *The Concept of Truth in Formalized Languages* (henceforth CTFL),¹ Tarski set himself the task of examining thoroughly under what conditions and by what methods it is possible to construct a satisfactory definition of the term 'true sentence', restricted to a given language, which satisfies the conditions of *formal correctness* and *materially adequacy*. As for formal correctness, Tarski required that the definition must be couched in clear terms that do not admit of any doubt and cause any methodological problems, and also that it must satisfy certain formal rules of definitions so that it will not be circular and inconsistent.² As for material adequacy, his fundamental insight was that a sentence such as

'Snow is White' is True iff Snow is White

Is a platitude that specifies the conditions under which a particular sentence of a language L is true, and because of that can be viewed as a partial definition of truth with respect to that particular L-sentence. And if a particular such biconditional defines truth with respect to a particular L-sentence, the totality of such biconditionals for L-sentences define truth for L.³

¹Work on this article was supported by the research grant n. P401/10/0146 of the Czech Science Foundation. It first appeared in Polish (Tarski, 1933), afterwards (revised and enlarged) in German (Tarski, 1935), its English translation coming last (Tarski, 1956). Throughout this article, I shall quote from the 2nd revised edition (1983). Other well-known papers of Tarski are (1936, 1944, 1969).

²In particular, as the notions of truth, denotation and satisfaction belong to one semantic family, hence raise similar worries (on account of semantic paradoxes or their allegedly metaphysical character), Tarski decided not to use any of them in the definition of truth, unless they can be reduced to purely non-semantic notions via explicit definitions.

³Such biconditionals, Tarski (1983: 187) maintained: '[...] explain in a precise way, in accordance with linguistic usage, the meaning of phrases of the form 'x is a true sentence', which occur in them. Not much more in principle is to be demanded of a general definition of phrases of the form 'x is a true sentence' than that it should satisfy the usual conditions of methodological correctness and include all partial definitions of this type as special cases; that it should be, so to speak, their logical product.'

According to Tarski's *semantic conception of truth* (SCT), biconditionals such as (1) hold and partially define the notion of truth, because they fix its application conditions (or adequate usage). More generally, biconditionals obtained from the schema *X is true iff p* (called 'T-schema') by replacing '*X*' by a perspicuous designator of a sentence of a language under consideration, and '*p*' by that sentence itself or by its translation, partially define the notion of truth for that language.¹ This is encapsulated in Convention T, which reads as follows:

A formally correct definition of the predicate 'true' in metalanguage ML will be called a materially adequate definition of truth for language L if it entails every T-biconditional that is an instance of the T-schema

X is true iff p

where '*X*' is a structural designator in ML of a sentence of L and '*p*' is either that sentence itself or its translation into ML.²

Truth Definitions for formalized Languages

Once the problem of truth definition has such a precise formulation, the question is for what languages it is possible to construct such truth definitions and the problem is to find a formal procedure for constructing a finite definition that implies all such biconditionals for (typically) infinitude of sentences of L. Tarski showed us how to define truth for a comprehensive group of properly regimented and fully interpreted languages, whose syntactic structure is exactly specified and are free of context-sensitive or ambiguous

¹Tarski (1969: 106).

²For the original formulation of Convention T see Tarski (1983: 187-188). Our formulation omits the second clause to the effect that every truth of L is a sentence of L. Some further comments are in order. [A] Although Convention T may appear to be general, as it stands it applies only when ML is English (or better, a well-behaved fragment of it), since instances of *X is true iff p* are English sentences. For a truth definition framed in another meta-language we have to reformulate the convention and use an appropriate version of T-schema (e.g., if ML is a fragment of German, we have to use *X ist wahr wnw p*). Convention T itself is formulated in a meta-meta-language, which may or may not contain ML. [B] If a definition meets Convention T, it is extensionally adequate – all and only the true sentences of L fall under the defined notion. For, any two definitions that satisfy Convention T are bound to be equivalent. But this does not mean that it also supplies a criterion of truth for L, that is, a method or test effectively deciding whether a given sentence of L is true or not. [C] Though the 'if' in Convention T suggests that it specifies only a sufficient condition of material adequacy, it was arguably intended to state both a necessary and sufficient condition of material adequacy, conceived of as a faithfulness to SCT (hence it does not explicitly disqualify alternative truth-definitions corresponding to different conceptions of truth). However, Tarski deemed it strange to uphold a conception of truth that is incompatible with SCT, since it presumably implies the denial of some T-biconditional, which denial amounts to an assertion of a sentence of the form '*p* is true iff not *p*'. And this is surely not a particularly attractive position.

expressions.¹ His strategy was to demonstrate how the method works for a particular formalized language belonging to a certain comprehensive group, and then to argue that the method can be extended to other language from that group.

For our purposes, it is not necessary to explain Tarski's method of truth-definition using his own example (the calculus of classes). It will be instructive to demonstrate how the method works for languages L_0 , L_1 and L_2 (or better, fragments of a full-blooded language) of increasing complexity, ending up with L_2 whose structure is analogous to the structure of those languages that belong to Tarski's group.

Let's start with L_0 , a formalized fragment of German containing three sentences 'Der Schnee ist weiss'; 'Das Grass ist grün' and 'Der Himmel ist blau', with their usual English interpretations (we assume that a well-behaved fragment of English is our meta-language). The following, list-like definition for L_0

(D1) For every sentence x of L_0 , x is true iff
($x =$ 'Der Schnee ist weiss' and snow is white)
or ($x =$ 'Das Grass ist grün and the grass is green)
or ($x =$ 'Der Himmel ist blau' and the sky is blue).

is perfectly adequate by Tarski's lights, because it is explicit, uses no undefined semantic terms, and all the partial definitions of truth for L (T-biconditionals for sentences of L) can be deduced from it, given certain obvious syntactical facts: 'Der Schnee ist weiss' \neq 'Das Grass ist grün'; 'Der Schnee ist weiss' \neq 'Der Himmel ist blau'; 'Das Grass ist grün' \neq 'Der Himmel ist blau'. To check this, let us substitute 'Der Schnee ist weiss' for ' x ' in the definition, thereby obtaining the following:

'Der Schnee ist weiss' is true iff ('Der Schnee ist weiss' = 'Der Schnee ist weiss' and snow is white) or ('Der Schnee ist weiss' = 'Das Grass ist grün' and the grass is green) or ('Der Schnee ist weiss' = 'Der Himmel ist blau' and the sky is blue).

Since we can eliminate the obviously true identity in the first disjunct as well as the obviously false second and third disjunct, what we are finally left with is the T-biconditional for 'Der Schnee ist weiss' — as wanted.

This method, though, does not work for languages containing more than finitely many sentences - unless we are prepared to allow infinite disjunctions

¹Tarski famously argued that in case of natural languages (or languages of approximately their expressive power) no such definition can be constructed. He had two principal reasons: (1) the inexact character of natural languages; (2) the universal character of natural languages that makes them *semantically closed* and hence inconsistent (a paradoxical sentence S stating its own untruth can be reconstructed in them that leads to a contradiction via logically plausible reasoning involving at its crucial step the T-biconditional for S). As we assume that the reader is familiar with this classic part of philosophy, we can save space here.

as acceptable definitions (as Tarski was not). But languages of theoretical interest have, as a rule, more complex structure, which forced Tarski to look for a more general method of truth-definition. Let us move to Tarskian truth-definitions for more complex languages. Let us expand L_0 to L_1 by adding the constructions ‘Es ist nicht der fall, dass’, ‘und’ and ‘oder’ (in their usual English interpretations), by iterated application of which an infinitude of compounded sentences can be formed from the three atomic sentences of L_0 . We can define ‘true’ for L_1 in a well-known recursive (inductive) manner (A and ‘ B ’ are taken to range over arbitrary sentences of L_1 and italicized complex expressions are used autonomously):

(D2) For every sentence s of L_1 , s is true iff ($s =$ ‘Der Schnee ist weiss’ and snow is white) or ($s =$ ‘Das Grass ist grün’ and the grass is green) or ($s =$ ‘Der Himmel ist blau’ and the sky is blue) or ($s =$ *Es ist nicht der fall dass A and A is not true*) or ($s = A$ und B and both A and B are true) or ($s = A$ or B and A is true or B is true)

(D2) is not an explicit definition, since the defined predicate appears in clauses fixing its application-conditions for compound sentences. With the help of some elementary set-theory at hand, however, it can be turned into an explicit definition, in which the original clauses are transformed so as to specify the conditions on membership in a certain set, call it ‘TRUE’:

(D3) For every sentence s of L_1 , s is true iff there is a set TRUE such that s belongs to TRUE, and for every sentence y of L_1 , y belongs to TRUE iff:

- (a) $y =$ ‘Der Schnee ist weiss’ and snow is white; or
- (b) $y =$ ‘Das Grass ist grün’ and the grass is green; or
- (c) $y =$ ‘Der Himmel istblau’ and the sky is blue; or
- (d) $y =$ *Es ist nicht der fall dass A and A does not belong to TRUE*; or
- (e) $y = A$ und B and both A and B belong to TRUE; or
- (f) $y = A$ or B and A belongs to TRUE or B belongs to TRUE.

It can be shown that, for any given sentence of L_1 , its T-biconditional is derivable from this definition, the material adequacy of the definition being thereby assured.

We have shown how to define the truth-predicate for L_1 so that T-criterion is satisfied, by fixing the truth-conditions of any complex sentence of L in terms of the truth-conditions of its less complex component sentences, ultimately in terms of the truth-conditions of simple sentences whose truth-conditions are fixed directly in their corresponding T-biconditionals. If, now, we expand L_1 by adding to it iterative constructions of a different kind, namely quantifiers, we obtain a potentially infinite number of complex sentences that we cannot deal with in this way, because their immediate constituents are no longer truth-evaluable sentences, but sentential functions, e.g.: ‘ x is blue’ or ‘ x

loves y ', ' x is a man and x loves y ', 'For all x , if x loves y , x is a man', etc. As sentential functions are neither true nor false, truth for the expanded language cannot be defined recursively, on the complexity of sentences of L . But Tarski had a good idea of how to fix the situation.¹ Because sentential functions can have more argument-places that are marked by different variables, which can be replaced by more terms or closed by prefixing to them more quantifiers, it is better to talk, generally, about satisfaction of n -argument sentential functions by ordered n -tuples of objects. He actually defined something more general: the relation x satisfies y , where x is an infinite sequence of objects (from the universe of discourse associated with L) and y a sentential function of L . This relation is the converse of the relation y is true of x , generalized so as to cover sentential functions with arbitrary (but finite) number of argument-places. To explain how this works, let us consider a regimented fragment of German, L_2 , that has no simple or complex terms (to keep the definition simple), and its only non-logical constants are: 'ist ein Mann', 'ist eine Frau', 'liebes' (all having their usual English interpretations). Atomic sentential functions of L_2 are formed by attaching one or two variables (taken from the infinite sequence: x_1, x_2, \dots, x_n) to its non-logical constants, while complex sentential functions and sentences of L_2 are formed from its atomic sentential functions via the operations of negation ' \neg ', conjunction ' \wedge ', disjunction ' \vee ', and universal quantification ' \forall ' (all with their usual interpretations).² Sentence of L_2 is defined as a limiting case of (0-argument) a sentential functions with no occurrence of a variable free in it (all occurrences of their variables being bound by a quantifier).³

The definition of the satisfaction relation for L_2 mimics the usual recursive definition of sentential function, proceeding by recursion on the complexity of a sentential function of L_2 : it is first specified under what conditions a given infinite sequence p of objects satisfies simple sentential functions of the form $x_i A$, and then the conditions are specified under which p satisfies complex sentential functions of the forms $A \wedge B$, $A \vee B$, $\forall x_i A$, in terms of the conditions

¹Tarski (1983: 189). Sentential functions are true or false only (a) relative to specific assignments of appropriate values to their free variables, or (b) relative to specific substitutions of denoting terms for their free variables. Now, if every object in the universe of discourse associated with a language L (e.g., a first-order language of arithmetic) is denoted by a term of L , we can well follow (b) and define the truth-conditions for quantified sentences of the form $\forall x_k A(x_k)$ in terms of the truth-conditions of sentences, in this way: *A sentence of the form $\forall x_k A(x_k)$ is true iff every sentence obtained by replacing the free variable in $A(x_k)$ by a (denoting) term t of L is true.* But, of course, L may not contain a name of every object in its universe of discourse (L may even have no names, as was the case with the language of the calculus of classes, for which Tarski provided his truth-definition).

²It should be clear that Tarskian definition cannot explain the meanings of logical or non-logical constants of the object-language, for which truth is defined. We can only define truth for such a language in Tarski's style if we already know their meanings (their translations or interpretations in the meta-language).

³Accordingly, sentential functions of L_2 have to be defined inductively.

under which p satisfies (or does not) less complex sentential functions that are their immediate components:¹

(D4) For every sentential function f of L_2 , an infinite sequence of objects p satisfies f iff

- (a) $f = 'x_k \text{ ist ein Mann}'$ and p_k is a man; or
- (b) $f = 'x_k \text{ ist eine Frau}'$ and p_k is a woman; or
- (c) $f = 'x_k \text{ liebes } x_l'$ and p_k loves p_l ; or
- (d) $f = \neg A$ and p does not satisfy A ; or
- (e) $f = A \wedge B$ and p satisfies A and p satisfies B ; or
- (f) $f = A \vee B$ and p satisfies A or p satisfies B ; or
- (g) $f = \forall x_k A$ and every sequence q that differs from p at most at the k -th place satisfies A

Now, whether or not p satisfies f depends only on those members of p that are paired with the free occurrences of variables of f , provided it has any. But, mind you, sentences are 0-argument sentential functions with no occurrences of variables free. Accordingly, whether or not a sentence is satisfied by p does not depend on what members of p are paired with their free variables. So, there are two possibilities: either a sentence is satisfied by all sequences, in which case it is true, or it is satisfied by no sequence, in which case it is not true. Truth for L is thus defined directly:

(D5) For every sentence s of L_2 , s is true iff for every y , y is a sequence $\rightarrow y$ satisfies x

Once again, the definition of satisfaction is implicit but it can be turned to an explicit one, provided that our meta-language has a sufficiently strong set-theory or higher order variables than L_2 , thereby eliminating the semantic notion of satisfaction in favor of non-semantic expressions of the metalanguage ((a) expressions of L or their translations, (b) set-theoretical expressions, (c) syntactical expressions). And all the T-biconditionals can then be derived from the explicit definition so that the material adequacy is assured.

Do Tarskian Truth definitions have any Semantic Import?

Although Tarski's method of truth-definition is considerably limited in its scope,² Davidson, Montague and others have persuaded many theorists that it

¹Henceforth I omit the following tedious qualifications: 'for some sentential function' or 'for some integer'. As usual, p_k is the k -th member of p . Since variables and objects in a sequence are ordered, every variable occurring in a sentential function is paired with exactly one object in the sequence, via its numerical index. Such pairing is then completely non-semantic.

²First, it is designed to apply only to a certain family of properly restricted extensional languages, without any suggestion whether, eventually how, Tarski's criteria and techniques might be supplemented or modified to cover syntactically richer languages, including natural

(or its model-theoretic extensions) provides at least a *Muster* that may guide constructions of more sophisticated semantic theories for richer languages, including substantial fragments of natural languages. In fact, it has even been claimed - although Tarski himself would not have gone so far - that plausible semantic theories for substantial fragments of natural languages should have *something like* Tarski-style truth-definitions as their basis, though modified or supplemented to accommodate features not present in languages with simple logical syntax. Donald Davidson went even so far as to suggest that we can use Tarski's method to give a compositional meaning for a given natural language L.¹

In view of this, the significance of Tarski's method of truth definition to formal semantics would seem to be beyond reasonable question. Yet, it has turned out that the question of significance of Tarski's method to semantics is a rather delicate matter. In fact, several philosophers argued that strictly Tarskian truth definitions are of no use as theories of meaning or even as semantic theories.

As regards Davidson's influential program, at least in its early form, one obviously pertinent observation is that we cannot expect from a Tarskian truth-definition for L that it tell us, via its satisfaction-conditions, truth-conditions or denotation-conditions, something revealing about the meanings of expressions and sentences of L, because we have to know in advance the meanings of non-logical and logical expressions in order to define truth for L in Tarski's style.² Thus, for instance, we cannot expect from the biconditional

'Der Schnee ist weiss' is True iff Snow is White

That it tell us both what the mentioned sentence means and what 'true' means. For it cannot serve as a partial explanation of the meaning of 'true' in Tarski style, unless we already understand the mentioned sentence and know that it means what the used sentence means. And it cannot serve as an

languages or languages approximating their complexity (one immediately thinks of so-called intensional or hyper-intensional constructions). Second, the method is designed to apply only to languages in which there is no context-sensitivity, no ambiguity or vacuous expressions. But natural languages, on which formal semanticists focus their attention, abound in such phenomena. Third, in view of semantic paradoxes, Tarski argued that no consistent definition or even theory of truth that meets his adequacy condition can be given for a language that contains or can express its own notion of truth (or satisfaction) unrestrictedly applying to all its sentences (predicates), provided that we assume that classical bivalent logic holds. But natural languages certainly do seem to contain such notions of truth (satisfaction), and so Tarski concluded that neither natural languages nor properly regimented languages that approximate them in expressive power can be given consistent truth-definitions in his style. Still, in spite of all these limitations, combined with Tarski's notorious skepticism towards natural languages,

¹Davidson (1984).

²This observation is commonly attributed to Dummett (cf. 1973), but Tarski was fully aware of it.

explanation of the meaning of the mentioned sentence, unless we already know what 'true' means in it.¹

Another objection, made famous by Putnam,² is that the logical or modal status of T-biconditionals for L differs from the logical or modal status of the biconditionals in which 'true' is replaced by its explicit Tarskian definiens. The informal T-biconditionals for L express empirical truths about sentences of L, contingent on their meanings (which, in turn, depend on practices, conventions or intentions of users of L). Had the usage or meanings of words and sentences of L changed in certain easily imaginable ways, the informal T-biconditionals for L could well change from true to false claims. Hence they express contingent empirical claims. But, as Putnam claimed, their Tarskian analogs are consequences of explicit definitions in a logico-mathematical axiomatic metatheory, augmented with axioms for syntax of L. On this basis he argued that Tarskian analogs are logical or mathematical, but certainly non-empirical, necessary truths that do not depend on the meanings of expressions and sentences of L, or on the facts about their use by speakers.

However, this line of argument can be blocked. Several people argued that Tarski takes L to be individuated in part by the fact that particular expressions belong to it that are equipped with particular meanings.³ So, if there is a single expression *e* whose meaning is different in L and L* respectively, then L and L* are different languages, even though they may be syntactically indistinguishable (comprised by exactly the same type-expressions). It follows that by adding a single expression to L or by changing the meaning of a single expression of L, we no longer have L but a different language L*. If we are imagining that a sentence or expression actually belonging to L could change its meaning contingently on its use by the community actually using L, we are imagining, strictly speaking, that a different, though closely related language L* would be used by the community. And once we individuate languages in this strict way, even informal T-biconditionals turn out necessary.⁴

¹In a similar vein, then, the recursive clause of the type *A und B is true iff A is true and B is true*, cannot explain in the same breath what 'true' means and what 'und' means. It follows that the project of defining truth for L in strictly Tarskian manner is incompatible with the project of a theory of meaning for L, as conceived of by Davidson and his followers. Meanwhile, Davidson came to concede this point, and in the mid-70s started to talk about reversing Tarski's strategy: instead of taking the meanings of L-expressions for granted and defining truth-in-L, he proposes to take truth for granted (as a primitive concept) and give the meanings via a recursive truth-theory (as opposed to definition) for L.

²See H. Putnam (1988, 1994 a, b).

³See Davies (1981), Garcia Carpintero (1996), Künne (2003).

⁴It should be noted that Putnam's argument could be blocked in a different way. We may, for instance, allow that L retains its identity even though its expressions change their meanings and semantic properties (contingently on their use by speakers of L), so that sentences that actually serve as true T-biconditionals for L can turn out false, other sentences serving as T-biconditionals for L. We may refuse to accept that Tarskian biconditionals as necessary truths derived from definitions added to a logico-mathematico-syntactical metatheory. Instead, we may argue that their modal status depends on the modal status of the definitions, and there is no hint that Tarski considered them as necessary, as opposed to materially true (rather, it is arguable that he preferred to avoid intensional notions). For lack of space I cannot deal with this interesting response here. But see Patterson (2008b).

A closely related but more serious objection is concerned with one specific aspect of Tarski's method of truth definition: whether simple or complex, implicit or explicit, Tarskian definitions rest, in one way or another, on list-like or enumerative definitions of semantic notions (or truth, satisfaction, or denotation). To explain what is at stake, let us recall the truth-definition we gave for L_0 . If we look at it, we see that there is nothing obviously *semantic* about it, even though it meets T-criterion and is thus faithful to Tarski's semantic conception of truth. It does not throw any light upon how the truth-conditions of sentences are determined by the semantic properties, including semantic relations objects, of their significant constituents and their manner of composition (it is no use to say that it relates sentences to extra-linguistic entities, facts or states of affairs, since it obviously does not do that). In light of this, the truth definition for L_2 certainly appears to be more illuminating of the semantic structure of L_2 than the two previous truth-definitions are of the semantic structures of L_0 and L_1 respectively. So, what this shows is that it is possible for a definition of truth for a language (at least for a simple language) to satisfy the demands of Tarski's semantic conception of truth, while not being *semantic* in any natural sense of that notion. That the truth predicate defined in this way does not have the right connections to semantic facts becomes transparent once we realize that if we did not understand German but had reliable information that the following informal T-biconditional is true

'Der Schnee ist Weiss' is True (in German) iff Snow is White

We would have at least some information about the meaning that 'Schnee ist weiss' has as a sentence of German, hence as a sentence of L_0 (assuming we understand the meta-language in general, and 'true' in particular). We could infer, for instance, that it does not mean that snow is not white, that snow is black and other things incompatible with the fact that snow is white. Now, it is standardly assumed that an explicitly defined predicate can always be replaced by the definiens without any loss. But when we replace 'true' in the informal T-biconditional for 'Der Schnee ist weiss', by its explicit Tarskian definiens what we get is this:

(*'Der Schnee ist weiss' = 'Der Schnee ist weiss' and snow is white*)
or (*'Der Schnee ist weiss' = 'Das Grass ist grün' and the grass is green*) or (*'Der Schnee ist weiss' = 'Der Himmel ist blau' and the sky is blue*) iff snow is white.

After performing admissible simplifications, this is equivalent to the following triviality

Snow is White iff Snow is White

Which does not state any information at all about the meaning or semantic properties of the sentence 'Der Schnee ist weiss'. Clearly, one can understand

what this list-like truth-definition states without knowing anything at all about the semantics of L_0 .

We have said that the truth definition for L_2 certainly appears more illuminating of the semantic structure of L_2 than the truth-definitions for L_1 is of the semantic structures of L_1 , and this in turn appears more illuminative than the truth definition for L_0 (because using recursion). But the appearances may be delusive. In fact, the critics argued, that essentially the same considerations apply, *mutatis mutandis*, to the truth definition for L_1 and L_2 . Especially if they are put in their explicit forms (and only these meet all Tarski's strictures), it becomes clear that one could understand what they state without knowing any semantic fact at all about L_1 or L_2 . As Etchemendy and Soames have pointed out,¹ when we replace in the T-biconditional for 'Der Schnee ist weiss' (as belonging to L_1), the predicate 'true' by its explicit Tarskian definiens, we get the following long claim:

There is a set TRUE of sentences of L_1 to which 'Der Schnee ist weiss' belongs such that...

It could be shown that after admissible simplifications of this claim we get something that, again, does not state any information at all about the meaning or semantic properties of the sentence 'Der Schnee ist weiss' in L_1 .² But, even at the intuitive level: we readily understand this claim, but unless we know that TRUE is the set of true sentences of L_1 (which is not stated in the definition), we cannot, on the basis of this claim, infer anything concerning the meaning that 'Der Schnee ist weiss' has as a sentence of L_1 .

Moving finally to the truth-definition for L_2 , we immediately observe that the part of it that deals with the satisfaction conditions of simple sentential functions is trivially list-like or enumerative:

p satisfies f iff ($f =$ 'x_k ist ein Mann' and p_k is a man)
or ($f =$ 'x_k ist eine Frau' and p_k is a woman)
or ($f =$ 'x_k liebes x_l' and p_k loves p_l)

But then essentially the same line of argument can be used to show that whole definition has no semantic import.

Where does the Semantic Import of Tarski's Method Lie?

It would seem that we must grant the critics that Tarski's method of truth definition has no semantic import. One cannot *stipulatively* introduce semantic notions for L via purely logico-syntactico-mathematical definitions, and at the same time expect that such definitions will state something relevant about the semantic properties and structure of L . Once satisfaction or denotation and, in

¹Etchemendy (1988: 56-57); Soames (1999: 102-105).

²Cf. Soames (1999: 104).

terms of them, truth is defined explicitly, there remain no semantic terms in their definiens, because they all disappear completely in favor of the terms of a semantics-free meta-theory. This might be all to the good, if one is after extensionally adequate definitions of semantic notions in terms of set-theory, logic, syntax and the vocabulary of the object-language, to which the definitions are relativized. This was arguably Tarski's aim, since he wanted in that way to show that restricted semantic notions are only as controversial as the apparatus in terms of which they are introduced. However, set-theoretical definitions, like those that are trivially list-like, can capture only the extensions of restricted semantic notions but cannot possibly analyze, explain or explicate them. The critics are right to rebut his loose claims to the effect that Tarskian definitions tell us all we can or need to know about the notion of truth. In view of this, Tarski's notorious claim that his truth definitions aim to catch hold of *the actual meaning* of our intuitive semantic notion of truth is unfortunate and misleading, to say the least.¹

On the other hand, we do not think that we can simply rest content with the conclusion that Tarski's method of truth definition has no semantic import. The point – which should be clear anyway but is often ignored by the critics – is that there is more to Tarski's method or conception of truth definition than particular formal definitions. Clearly, its heart is Convention T, stating the material adequacy condition on a satisfactory definition of truth for a given language, and the formal technique is, in general case, recursive characterization of satisfaction (or denotation) based on the logico-syntactic complexity. In tandem, these two aspects indicate to us where the semantic import of Tarski's method lies.

It is surely an inherent part of Tarski's method that the truth definition for L is meant to be a materially adequate definition of truth for L, faithful to one powerful intuition about our ordinary notion of truth. This intention takes the form of the requirement of its implying all T-biconditionals for the language under study. The fact that it has such consequences confers upon it the status of the truth definition for L, for then the claim that all and only true sentences of L satisfy the definition holds. Consequently, Tarskian truth definition cannot be a purely stipulative definition, although it cannot reasonably be construed as a meaning or concept giving definition either – as providing a meaning analysis or explication of the ordinary notion of truth (as the foregoing arguments of the critics showed).²

The whole semantic import of a Tarskian truth definition for L, as based on the recursive definitions of satisfaction for predicates (or denotation for terms) of L, rests on the claim that the definition is a materially adequate definition of the truth for L. True, in making this claim we must appeal to our ordinary notion of truth. But there is no circularity involved, because that claim itself is no part of the definition, but a meta-claim about our success in fulfilling what was our intention. Note that Tarski himself did not hesitate to appeal to the

¹Tarski (1944: 360).

²The fact that it is intended to be materially adequate also shows that it is not divorced from meaning but depends on it (viz. Convention T and its appeal to the notion of translation).

notion of truth in his version of Convention T: ‘A formally correct definition of the symbol ‘Tr’, formulated in the metalanguage will be called an adequate definition of *truth* if ...’.¹ Once we know what it takes for a definition of truth to be materially adequate and that the proposed definition of truth for L is materially adequate, we can read it as containing relevant information about the semantic properties of L, based on its compositional structure.

Now, as Etchemendy, Davidson and Heck persuasively argued,² the claim that Tarskian truth definition for a full-blooded quantificational L is materially adequate (satisfies Convention T) makes the definitions in a sense equivalent to an axiomatic theory of truth for L, whose axioms mimic the clauses of the recursive truth definition (*viz.* e.g. (D4) on which (D5) is based), with semantic notions construed as primitives. For various reasons that I shall not rehearse here, the question of whether axiomatic theories along these lines can serve as empirical theories of understanding/interpretation, as Davidson claimed, is rather controversial.³ But, at the very least, such theories do seem to shed some light upon the semantic structure of quantificational languages: how the truth conditions of sentences systematically depend on the truth-relevant semantic properties of their significant parts based on their logico-syntactic structure (thus revealing the inferential structure of quantificational languages).

In our view, Tarski was well aware that his method of truth definition had this dimension (on which also his later model-theoretic views rest). He said that in order to capture the truth conditions for infinitude of sentences of a reasonably rich (quantificational) formalized language it is *natural* to proceed through a recursive characterization of the satisfaction conditions for complex predicates in terms of the satisfaction conditions of less complex predicates that are their immediate constituents (if L has complex terms, a recursive characterization of the denotation conditions being added).⁴ He deemed it natural, we dare say, because it is a rather intuitive thought that semantic properties of complex expressions, *a fortiori* truth conditions of sentences, systematically depend on semantic properties of their significant constituents.⁵ When we are concerned with languages worth of that name (and not with oversimplified fragments like L₀ and L₁), a satisfactory characterization of

¹Tarski (1983: 187-188).

²Etchemendy (1988), Davidson (1990), Heck (1998).

³Davidson (1990).

⁴Cf. this statement of Tarski (1944: 345): ‘[...] it turns out that the simplest and the most natural way of obtaining an exact definition of truth is one which involves the use of other semantic notions, e.g. the notion of satisfaction [...]’.

⁵Logicians have always honoured intuitions such as the following: a sentence of the form *name+predicate* is true iff the predicate is true of what the name denotes (alternatively: ...iff what the name denotes has the property that the predicate denotes). Such intuitions were elaborated and generalized in various ways, but the essential idea remained: truth or falsity of sentences (of at least certain forms) depends on the semantic properties of their significant syntactic parts, so that it is possible to specify the condition under which a sentence (of an appropriate form) is true in terms of the semantic properties of its parts. This idea, in the air at least since Plato’s *Sophist*, was in modern logic worked out by Frege (his compositionality principles based on a sort of categorial grammar). It would be very uncharitable to assume that a prominent member of Lvov-Warsaw school was unaware of it.

truth conditions for sentences of a language like the one we gave for L_2 is naturally going to be framed in terms of the relations of satisfaction or denotation (or something of the sort) pairing expressions and objects; and, typically, such relations will be defined recursively on the logico-syntactic structure of expressions. Such a definition then displays the contribution of that structure to the truth-conditions of sentences of L_2 , and provides in this way also the basis for a precise account of logical consequence. And that is no mean achievement. This, in our view, is the real semantic import of Tarski's method of truth definition, in which also its philosophical importance largely consists.

Bibliography

- Davies, M. (1981). *Meaning, Quantification and Necessity*. London: Routledge.
- Davidson, D. (1984a). *Inquiries into the Truth and Interpretation*. Oxford: Oxford University Press.
- (1990). 'The Structure and Content of Truth.' *The Journal of Philosophy* (87): 279–328.
- Dummett, M. (1973). *Frege: The Philosophy of Language*. London: Duckworth.
- Etchemendy, J. (1988). 'Tarski on Truth and Logical Consequence.' *The Journal of Symbolic Logic* (53): 51–79.
- Field, H. (1972). Tarski's Theory of Truth.' *The Journal of Philosophy* (69): 347–75.
- Garcia Carpintero, M. (1996). 'What is a Tarskian Definition of Truth?' *Philosophical Studies* (82): 113–44.
- Heck, R. (1997). 'Tarski, Truth and Semantics.' *The Philosophical Review* (106): 533–54.
- Künne, W. (2003). *Conceptions of Truth*. Oxford: Oxford University Press.
- Montague, R. (1974). *Formal Philosophy. Selected Papers of Richard Montague*. New Haven: Yale University Press.
- Patterson, D. (ed.) (2008a). *New Essays on Tarski*. Oxford: Oxford University Press.
- (2008b). 'Tarski Conception of Meaning.' In: Patterson (2008a), pp. 157–191.
- Putnam, H. (1988). *Representation and Reality*: Massachusetts: MIT Press.
- (1994a). *Words and Life*. Ed. J. Conant, Harvard: Harvard University Press.
- (1994b). 'On Truth.' In: Putnam (1994a), pp. 315–29.
- (1994c). 'Comparison of Something with Something Else.' In: Putnam (1994a), pp. 330–50.
- Soames, S. (1984). 'What is a Theory of Truth?' *The Journal of Philosophy* (81): 411–29.
- (1999). *Understanding Truth*. Oxford: Oxford University Press.
- Tarski, A. (1933). 'Pojęcie prawdy w językach nauk dedukcyjnych.' *Prace Towarzystwa Naukowego Warszawskiego, Wydział III*, no. 34.
- (1935). 'Der Wahrheitsbegriff in den formalisierten Sprachen.' *Studia Philosophica* (1): 261–405. English translation in: Tarski (1956, 1983a), pp. 152–278.
- (1936). 'Grundlegung der wissenschaftlichen Semantik.' In: *Actes du Congrès International de Philosophie Scientifique*, vol. III, Paris, pp. 1–8.
- (1944). 'The Semantic Conception of Truth and Foundations of Semantics.' In: *Philosophy and Phenomenological Research* (4): 13–47.

An Anthology of Philosophical Studies

- (1956). *Logic, Semantics, Metamathematics*. Ed. J. Woodger, Oxford: Oxford University Press. 2nd edition Tarski (1983a).
- (1969). 'Truth and Proof.' *Scientific American* (220): 63–77.
- (1983). *Logic, Semantics, Metamathematics, papers from 1923 to 1938*. Ed. John Corcoran, Indianapolis: Hackett Publishing Company.

CHAPTER SIXTEEN

Being as the Set of All Things

Donald V. Poochigian

Presumably shaped by the Milesians, Classical Greek philosophy identifies being as the universal primitive, an inexplicable simple constituting all else. Illustrative, “Aristotle regarded metaphysics as the study of being, or of ‘being *qua* (as) being’ (that is, being ‘as such’). Since there is really not very much to say about being *qua* being, even Aristotle ended up talking about many other less general (i.e., less basic) categories.”¹ Overlooked is problematic character of “categories.”

Anaximander identifies being as content; Anaximenes identifies being as content and form. Since same content can appear as whole (condensed) or parts (rarefied), form is not content. Not content, it is ontologically distinct. “Being *qua* being” is now complex, not simple.

Typically resolution is sought by appeal to the simplicity criterion of Occam’s razor. Ambiguous, though, simplicity is only apparent, complex as such. Appearance is contradictory, concurrently one and many, fused whole and diffused parts. Part immediately contiguous to part, all uninterruptedly flow into one another, even space a visually and tactilely perceivable something. Inconsistency is resolved by appeal to reality. Qualitative experience can be real or unreal, real when manifesting an unapparent state of being. Unapparent, reality is unobservable, imaginary as so. Introduced is abstraction, a qualitatively unimaginable entity. As such, it is real.

What is designated “real” encompasses sensate quality and abstraction; what is designated “modal” encompasses imaginary quality. Modes constitute intentional constituents of a qualitative set. Sensate quality is “real” because shared by at least most human beings, and imaginary quality is “modal” because not shared by at least most human beings. Intensional criterion of set membership can be real or imaginary quality, identity of a predictive entity verifying a scientific theory being imaginary. Yet unobserved, such an entity only can be identified imaginatively since, “we can very freely *call to mind* . . . propositional . . . content,”² mind being “an *active* capacity, in that we can, within limits, employ it successfully at will.”³

This is so even if identity is definitional rather than analogical, definition being imaginative since unobservable. A limited conjunctive sequencing of properties modeling an incomplete class member, definition cannot

¹William James Earle, *Introduction to Philosophy* (New York: McGraw-Hill, Inc., 1992), 91.

²Robert Audi, *Epistemology: a contemporary introduction to the theory of knowledge*, 2nd ed. (New York: Routledge, 2003), 91.

³*Ibid.*, 119.

accommodate an additional exclusionary property. Such is identifiable only by comparison with an analogical archetype, this latter composing an unlimited conjunctive sequencing of properties modeling a complete class member.

Distinction between the two worlds of appearance and reality is itself liable to the accusation of complexity. Classical positivism is interpretable as rejecting reality for an ever more complex appearance. Law is conjoined with law in limitless succession. Logical positivism is interpretable as accepting reality for an ever more simple appearance. Law is conjoined with law in limited succession.

To accommodate the paradoxical character of appearance, however, reality itself is paradoxical. It can be unitary to accommodate appearance as fused one, as with Baruch Spinoza, or it can be plural to accommodate appearance as diffused many, as with Gottfried Wilhelm Leibniz. Manifest is ambiguity of Occam's simplicity criterion.

Property

Conditional of everything, being might be thought a property. Since at least Immanuel Kant, however, it has been commonly assumed not.¹ How can being be a property of being, something be a property of itself? Being is not a property if it is trans-environmental; quality is a property if it is not trans-environmental. Presumed is being is universal but property is relative. Something is or is not in all environments, but if it is, can have different properties in different environments. Yet as one author expresses, "I find this criticism inconclusive. I don't see why existence isn't a property of existing things."²

¹Kant's diagnosis of the [ontological] argument's chief fault can be stated as follows:

Now suppose that there are two classical scholars, one of whom thinks that Homer existed and one of whom thinks Homer was legendary (the two great epics that are supposedly his compositions having been pieced together over a long period from the work of many anonymous poets). It would be wrong—in fact, it would be absurd—to describe the disagreement of these two scholars by saying that the former thinks that someone had the collection of properties H and, in addition, the property of existence while the latter agrees that someone had the collection of properties H but goes on to assert that this person lacked the property of existence. No, it's just that the former scholar thinks that someone had all of (or at least most of) the properties in the set H and that the latter thinks that no one had all of (or even very many of) them. This case illustrates the sense in which existence is not a property.

Peter van Inwagen, *Metaphysics* (Boulder, Colorado: Westview Press, 1993), 77.

²So Kant's claim is that Anselm made the mistake of treating existence as a property of existing things. I find this criticism inconclusive. I don't see why existence isn't a property of existing things. Granted, it may sound funny to talk about a property that everything under the sun possesses; most properties apply to some things but not to others. And maybe it is true that 'Tigers exist' can be paraphrased as 'The concept of tigerhood is exemplified.' However, these facts don't show that existence isn't a property of existing things.

Anonymous, *Questions in Philosophy*, blind review manuscript (Upper Saddle River, New Jersey: Prentice Hall, 2003), 88.

Kant's challenge conclusively proves being is not a property only if being is universal. If being is everything, it is the set of everything. But the set of everything is self-contradictory because containing more than itself, both itself and everything else. Not containing more than itself, either it contains everything else and not itself, or it contains itself and nothing else. Both containing and not containing itself, the set of everything is paradoxical. As the set of everything, being is paradoxical. Thus, everything having being, how is being distinct from the set of everything when the set of everything is paradoxical, either not the set of everything if not containing itself to avoid the contradiction of containing more than itself (every other set and the set of itself), or containing itself and nothing else to avoid the contradiction of containing more than itself (the set of itself and every other set)?

Being is not a property because for it to be so requires it being constituent of something already having being, begging the question by presupposing what is to be determined. But being can be a property if occurring concurrently with whatever has it as a constituent. This possibility requires being to exist *qua* being, independently of anything else. If this can be, then at least it is possible for something other than being itself to occur concurrently with it. And it is possible for being to exist *qua* being, it being possible to imagine simple existence, which is identity as distinct from everything else, and nothing more. Identity as not distinct from everything else is complex being, and when occurring concurrently with something other than itself, confers being on that thing.

Unconsidered is being can occur in different states. Doing so, the set of all states is universal, when a distinct state is particular. Same thing can be a property of itself when property and substance are different instances and/or aspects of the same thing—front and back spatially, now and then temporally, criterion and instance self-identifiably. Similarly, same thing can be a property of itself when property and substance are different conditional modal states of the same thing.

Something can exist in one environment such as imagination, and not in another environment such as sensation. A being existing in no environment is literally unimaginable because it would not exist in at least the domain of imagination. Thus, being is particular as well as universal. As particular, it is a property if quality is a property. But of what is being a property if it is not trans-environmental? It can be identity of an aspect or instance in an environment with an instance or aspect in another environment. Or it can be identity of an aspect or instance—whether modal, temporal, or spatial—with the set of all instances and aspects of something.

As these examples indicate, a property identifies a member of one set with a member of another set. What, though, is identity? It is membership in a common set. As so, George Boole understands it as membership in the subset of common members of different sets. But every member of the set of you yesterday is identified with every member of the set of you today, when no member of either set is common to (contained in or a member of) the other set.

Resolution is by making the two sets subsets of an encompassing common set. And this occurs by employing the logical operation of implication. All disjunctives of different sets (ab), real or imagined, are resolved inclusively (both ab). Doing this fuses the two diffused sets a and b into the one set AB . Now every member of what become subsets a and b are members of the common set AB , when membership in a common set is identity.

Specification occurring by observable designation, whether behavioral or verbal, when observable designation composes related parts, and related parts are observationally indistinguishable from unrelated parts, unclear is how “particulars in the relevant resemblance class are meant to resemble the paradigm is appropriately specified.”¹ Appeal can be to another observable designation, but how is it known the elements of this observable designation are related parts, etc.? Related and unrelated parts being observationally indistinguishable, what identifies the parts as related or unrelated must be unobservable.

It might be argued being is not a “property” because for it to be so requires it being constituent of something already having being, begging the question by presupposing what is to be determined. Concluding, “the sign for a function already contains the prototype of its argument, and it cannot contain itself,” Ludwig Wittgenstein makes this point.² Asserting “distinction . . . between a subject matter under study and discourse about the subject matter,” David Hilbert makes a similar differentiation.³ Meaning distinguished by a meta-function intends the function, different meta-functions of a function having different meanings not only than the function, but one another.⁴

To think a set can contain itself and more commits “a category mistake” representing “facts . . . as if they belonged to one logical type or category (or range of types or categories), when they actually belong to another.”⁵ As non-self-exemplifying, something being itself cannot be a property of itself without engendering the contradiction of being concurrently not greater and greater than itself. But as not non-self-exemplifying, something being constituent of itself can be a predicate of something else without engendering the contradiction of being concurrently not greater and greater than itself.

Here being can be a property if occurring concurrently with whatever has it as a constituent. This possibility requires being to exist qua being,

¹ . . . it may be objected that resemblance between particulars is always resemblance *in some respect*—for instance, in respect of size, or shape, or colour—and that unless the respect in which particulars in the relevant resemblance class are meant to resemble the paradigm is appropriately specified, an appropriate resemblance class will not be specified by the proposed method. E. J. Lowe, *A Survey of Metaphysics* (New York: Oxford University Press, 2002), 356.

²Ludwig Wittgenstein, *Tractatus Logico-Philosophicus* (Atlantic Highlands: New Jersey, 1974), nt. 3.333, 17.

³Ernest Nagel and James R. Newman, *Gödel's Proof* (New York: New York University Press, 1986), 31.

⁴Ludwig Wittgenstein, *Philosophical Investigations* (New York: The Macmillan Company, 1968), nt. 333, 17.

⁵Gilbert Ryle, *The Concept of Mind* (New York: Barnes & Noble Books, 1949), 16.

independently of anything else. If this can be, then at least it is possible for something other than being itself to occur concurrently with it. And it is possible for being to exist qua being, it being possible to imagine simple existence, which is identity as distinct from everything else, and nothing more. Identity as not distinct from everything else is complex being, and when occurring concurrently with something other than itself, confers being on that thing.

Quality

Being is emergent if basic and caused if not basic, simple as basic and complex as not basic, irreducible as simple and reducible as complex. Experience composes quality and abstraction. Quality is extended in time and/or space; abstraction is extended in neither, but can occur in both. Abstract awareness can occur over time and in a space. But quality is different at different times and in different spaces, and abstraction is not.

Indicated is time and space are constitutive of quality, a defining condition, and not of abstraction. Time and space are endlessly divisible. They being so, quality is endlessly divisible, intrinsically complex as such, not simple. Not being so, abstraction is simple in this respect. An abstraction can be simple, this is nothing else, or complex, this and this are nothing else. As complex, it is not simple being as parts and not whole, and is simple being as whole and not parts—this complex and nothing else. Distinguishing, simple abstraction constitutes identity. Simple abstraction constituting being, identity constitutes being. Abstraction constituting being, can quality have being?

Every experiential quality can be emergent, deduced from an infinite sequence, or induced into an infinite sequence. Emergence, deduction, and induction is identity. Different identity is ambiguity. Resolution of ambiguity is nominal. Qualitative identity is resolution of ambiguity. Therefore identity of an experiential quality is nominal.

Every conscious occurrence can be conjoined with another conscious occurrence in a sequence. Every sequence can be negated. Every negated sequence is a different sequence. Therefore, there are an infinite number of sequences. Every sequence can be conjoined with another sequence. Every conjunction of sequences is a different sequence. There are an endless number of sequences. Therefore, there are an endless number of conjunctive sequences.

Every quality is extended in time and/or space. Time and space are endlessly divisible. Therefore, every quality is endlessly divisible. Every division of a quality is a different quality. Therefore, there are a limitless number of qualities.

Every argument of a quality is an identity of the quality. There are unlimited deductive and inductive arguments of a quality. Every deductive and inductive argument of a quality is an identity of the quality. Every identity of a quality is real. Therefore, no identity of a quality is more real than another identity of the quality.

Qualitative self-evidence, reidentity, and similarity are self-exemplifying, tautology self-identifying as so and analogy other-identifying as so. *Haecceity* is characteristic of abstraction and qualitative identity, self-evidence as so, not self-identity, there being no variation to reidentify. *Tautology* is characteristic of qualitative reidentity, self-identity as so, not self-evidence, there being variation to reidentify. *Analogy* is characteristic of qualitative coidentity, similarity as so, neither self-evidence nor self-identity. Tautology identifies unambiguous members of a set; analogy identifies ambiguous members of a set. Unambiguous members of a set not being limits, tautology identifies non-limiting members of a set. Ambiguous members of a set being limits, analogy identifies limiting members of a set.

Arguably Robert Audi distinguishes identity with,
We can by and large introspect *at will* - roughly, just by
(sufficiently) wanting to - though we may also do it quite
spontaneously. . . . Granted, some content - like sensations of pain -
comes into consciousness uninvited; still, we can very freely *call to*
mind both propositional and imagistic content.¹

This is reasserted with, “Reason - conceived roughly as our mental capacity of understanding, especially in conceptual reflection or in inference - . . . is an *active* capacity, in that we can, within limits, employ it successfully at will.”² Reaffirming “We can . . . introspect . . . quite spontaneously,” with “some content . . . com[ing] into consciousness uninvited,” is the Japanese *waka* poem, “Thoughts flare up noiselessly, even the glow-worm appears without a sound.”³ To be emphasized in these passages is “*at will*,” autonomous mind being a precondition of reason, when “Logic alone does not guide such decisions.”

Abstraction

Constitutive of qualitative identity is abstraction, quality having being when informed by abstraction. Constant in any time or space when inconstancy in any time or space is extension in time or space, abstraction is extended in neither time nor space. Any abstraction in a different time or space is the same abstraction, abstraction not being casuistic. Inconstant in any time or space, quality is extended in time and space. Any quality in a different time or space is a different quality, quality being casuistic. Therefore no compounding of abstractions can constitute quality, and no deducting of qualities can constitute abstraction.

¹Audi, *Epistemology*, 90-91. See also, “Reason - conceived roughly as our mental capacity of understanding, especially in conceptual reflection or in inference - . . . is an *active* capacity, in that we can, within limits, employ it successfully at will.” *Ibid.*, 119.

²*Ibid.*, 119.

³Gabriele Fahr-Becker, *Japanese Prints* (New York: Barnes & Noble Books, 2003) 53.

Quality presents an “**irrational number** . . . : a number that can be expressed as an infinite decimal with no set of consecutive digits repeating itself indefinitely and that cannot be expressed as a quotient of two integers.”¹ Abstraction presents a “**rational number**,” that can be expressed as “a complete entity.”² Observationally indistinguishable as one or many, and if many, an endless number of many, quality is distinguishable as any by abstraction. Either identifiable integrally with quality or identifiable separately from it, abstraction is constituent of quality, infusing every aspect of it.

All that distinguishes different understandings of quality is the sense of identity. Only identifiable within experience are quality and abstraction, quality ontologically distinguished by extension in time and/or space, and abstraction by extension in neither time nor space. Quality encompasses real and imaginary objects. It is epistemologically distinguished by *sensation* of real objects, and *sensate-like* experience of imaginary objects. Abstraction is epistemologically distinguished by *sense*. Thus, there can be a sensate or sensate-like experience of quality, but not a sense of quality. Alternatively, there can be a sense of abstraction, but not a sensate or sensate-like experience of abstraction.

Quality as one or many, and if many, how many, being sensually indistinguishable, and alterable under continuous observation, it is distinguishable as any only by sense of identity. “Sense” ontologically distinguishing abstraction, identity is an abstraction. Any constituent of a quality constituting the same quality, identifying abstraction “infuses” a distinguished quality. Thus to be identifiable is “sense,” and unidentifiable is “nonsense.”

Being this and not anything else, simple identity is indivisible. As such, it occurs in neither time nor space, time and space being limitlessly divisible into endlessly smaller moments and areas. Occurring in neither time nor space, abstractions are limited because indivisible. Being limited, abstractions can be completely fused or diffused. Whichever abstractions are, there is no unresolved disjunctive of them.

Abstractions are simple identities, and disjunctives of simple identities are unimaginable. Resolution of disjunctives being relation, abstractions cannot be related. Rather than transitioning through disjunctive resolutions of states of being, abstractions emerge in states of being. This is because they occur in imagination rather than sensation. Experience appears as if uninterruptedly moving from world to world, different worlds not sharing content and/or form. Imagination and sensation are such different worlds. Abstraction spontaneously appears and disappears in imagination, rather than transform as in sensation.

¹“**irrational number**,” *Merriam Webster’s Collegiate Dictionary, tenth edition*, eds. Frederick C. Mish, et. al. (Springfield, Massachusetts: Merriam-Webster, Incorporated, 1996) 619.

²“**rational number** . . . : an integer or the quotient of an integer divided by a nonzero integer.” [“**rational number**,” *ibid.*, 969.] “**integer** . . . **1** : any of the natural number, the negatives of these numbers, or zero **2** : a complete entity.” [“**integer**,” *ibid.*, 607.] “**natural number** . . . : the number 1 or any number . . . obtained by adding 1 to it one or more times : a positive integer.” [“**natural number**,” *ibid.*, 774.] “**whole number** . . . : any of the set of nonnegative integers; *also* : INTEGER.” [“**whole number**,” *ibid.*, 1351.]

And although quality transforms in imagination, it need not in the way it transforms in sensation.

There being no disjunctive of such an abstraction, its identity is determinate absolutely. This is to be logically complete, and rational as so. Thus, abstract representation of quality and qualitative representation of abstraction are both necessarily logically incomplete. So it is a limited sequence is “sense,” and an unlimited sequence is “nonsense.” Abstraction being rational and quality being irrational, the problem of the one and many can be resolved abstractly, but cannot be resolved qualitatively.

Quality being limitlessly divisible, a field is an abstraction. As so, it is indivisible. A field composes simple or indistinguishable identities. A disjunctive of simple or indistinguishable identities is unimaginable. A field is not a relationship, then, and so neither sequence nor set. Being without relationship, it is indivisible. As so, it is not a set, which constitutes related objects. Not a set, it constitutes an urelement or “*urelements*, objects in our universe which are not sets or classes and which have no members.”¹

Being indivisible, there is no algorithmic operator by which a field is cleaved. Something always being between any other two things, rather than nothing, no purchase can be gained by which to prize the field apart. A field is cleaved by identifying autonomous elements as constituent of it. Set reduction is an iterative sequence employing the operation of identifying autonomous elements as constituent of it in a descending order of ever less encompassing subsets. Such an operation is possible ontologically because a field exists only in imagination. Being so, it is as possible to imagine constituents of a field as it is possible to imagine a field. Effectively imagination is the operator by which a field is prized apart. Separation of a field cannot be represented by an operational algorithm because imagination is primitive.

The Set of all Things

Being has been assumed indefinable because whatever the content of what is, constituent of it is being. Primitive as such, being is,

*The Unlimited . . . the first-principle of things that are. It is that from which the coming-to-be [of things and qualities] takes place, and it is that into which they return when they perish.*²

As with William James’ “stream of consciousness,” being presents,

¹Jon Barwise and Lawrence Moss, *Vicious Circles: On the Mathematics of Non-Wellfounded Phenomena* (Stanford, California: Center for the Study of Language and Information, 1996) 21. See also Roger Bishop Jones, “urelement,” *The Factasia Glossary*, <http://www.rbjones.com/rbjpub/pholos/glossary/u.htm>.

²Anaximander, “Fragment,” in *The Presocratics*, ed. Philip Wheelwright (New York: The Odyssey Press, The Bobbs-Merrill Company, Inc. Publishers, 1966) 54.

nothing jointed; it flows. . . . But now there appears . . . a kind of jointing and separateness among the parts, of which I refer to the breaks that are produced by sudden *contrasts in the quality* of the successive segments of the stream of thought.”¹

Like clay, being occurs only in its manifestations. Clay, independent of manifestation, even as a lump, is unknowable, when all pots are clay. Similarly, being, independent of manifestation, is unknowable, when all things are being. Fundamental, being is simple. Simple, it is irreducible. Irreducible, it is indefinable.

Unknowable independently of its manifestations, however, it is knowable dependently of its manifestations. It is thus as the set of all things. This when,

Classes and concepts may, however, also be conceived as real objects, namely classes as ‘pluralities of things’ or as structures consisting of a plurality of things and concepts as the properties and relations of things existing independently of our definitions and constructions.

It seems to me that the assumption of such objects is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence.²

As the set of all things, being composes one and many. The process of reformulating the set of all things generates its power set, iteratively identifying elements between elements within the limits of the power set. Relevantly, neither there is a subset, nor there is no subset, between any two subsets, is inconsistent with the power set.

Separating integrated elements by identifying what is between them until they are diffused generates discreteness. Integrating separated elements by identifying what is between them until they are fused generates denseness. Discreteness is distinction from every other thing and no extension, having number and no form. Denseness is distinction from every other thing and extension, having number and form.

Composing one and many, being is definable in terms of itself, alternately as one and many. Considering a commutative ring with unity,³ exhibited is a function whereby constituents reciprocally fuse into the same indistinguishable whole by repetitive iterative application of the “+” conjunctive function in any sequential order, and diffuse into the same distinguishable parts by repetitive

¹William James, “The Stream of Thought,” *Principles of Psychology*, in *Pragmatism: The Classic Writings*, ed. H. Standish Thayer (New York: Mentor, 1970) 142.

²Kurt Gödel, “Russell’s Mathematical Logic,” in *The Philosophy of Bertrand Russell*, ed. Paul Arthur Schilpp (Harper & Row: New York, 1963) 137.

³“A ring is commutative if the multiplication operation is commutative.” Eric W. Weisstein, “Commutative Ring,” From *MathWorld* — A Wolfram Web Resource, last updated: Wed Aug 26 2009. <http://mathworld.wolfram.com/CommutativeRing.html>. “Two elements x and y of a set S are said to be commutative under a binary operation $*$ if they satisfy $x*y=y*x$,” *ibid.*, last updated: Wed Aug 26 2009. <http://mathworld.wolfram.com/Commutative.html>.

iterative application of the “.” disjunctive function in any sequential order.¹ Constituents as fused whole are an object, and as diffused parts are objects.

Identified is conjunction, relation of identifiable particulars. Relation is a continuum between particulars. How is a particular identifiable within a continuum? An identifiable constituent of a continuum is represented reciprocally. Fusion of constituents renders a particular indistinguishable from the continuum. Diffusion of constituents renders a particular distinguishable from the continuum. Cyclic transitive marginal conversion of constituents from fusion to diffusion and diffusion to fusion renders a particular distinguishable and indistinguishable from the continuum. Thus, being is the alternating “contraction and condensation, and . . . distension and rarefaction” of the set of all things.²

Being something, resolved is the paradox of the set of all things containing and not containing itself. A non-empty set can contain or not contain itself. To contain itself is not to be constituent among other constituents of itself, for the set would be concurrently greater and not greater than itself. Self-contradictory, the set would be unidentifiable, indeterminate as to whether the greater or not greater. A non-empty set does not contain itself when its elements are unrelated, and does contain itself when its elements are related.

As parts a set does not contain itself; as whole it does contain itself. A thing as parts inductively identifies itself as whole, and the thing as whole deductively identifies itself as parts. Although an endless circle, such proof is rational, not vicious, equivalent to the mathematical criterion of a rational number as a “set of consecutive digits repeating itself indefinitely.”³ This is unlike an infinite regress which is vicious, not rational, because not a “set of consecutive digits repeating itself indefinitely.”

¹Ian Stewart, *Concepts of Modern Mathematics* (New York: Dover Publications, Inc., 1995) 78. For a more extensive definition of a commutative ring with unity see *Ibid.*, 76-78.

²Plutarch *Moralia* 947F, in *The Presocratics*, ed. Wheelwright, 62.

³“irrational number,” Mish, 619.

CHAPTER SEVENTEEN

The Self's Survival and its Freedom

Andrew Ward

Two contrasting theories of our survival or identity over time are prominent in the contemporary literature. On the one hand, there is the theory known as *psychological reductionism* or *the psychological reductionist approach*¹ and, on the other hand, the theory known as *animalism* or *the biological approach*². In the first two parts of the paper, I argue that neither theory has succeeded in disproving that our survival resides in the persistence of a separately existing subject of experiences. I shall contend, on the contrary, that the concept of such a subject not only fits well with the everyday belief that we have of ourselves, but that its central core can actually be justified. In the third part, I maintain that this justified concept of a persisting subject of experiences enables us to defend a robust sense of freedom of the will: a sense that constitutes a perceptible enrichment of the self's autonomy as this is currently conceived by compatibilist theories of freedom.

I.

According to the psychological reductionist, our survival, or at least what *matters* for our survival (what has moral or rational significance in our concern to survive), just consists in a succession of experiences, linked together by the causal relations necessary for the retention of memories, personality traits, and intentions. The conviction, which the reductionist acknowledges that we do naturally have, namely that our continued existence essentially resides in the persistence of a self or subject of experiences – understood as a continuant that is *separable* from any given succession of causally related experiences and/or

¹Two of its leading exponents are Sydney Shoemaker and Derek Parfit. See S. Shoemaker 'Personal Identity: A Materialist Account', in S. Shoemaker and R. Swinburne *Personal Identity* (Blackwell, Oxford: 1984) and *The First Person Perspective and Other Essays* (Cambridge University Press, Cambridge: 1996); and D. Parfit 'Personal Identity', *Philosophical Review* Vol. 80 (1971), *Reasons and Persons* (Part III) (Oxford University Press, Oxford: 1984), 'The Unimportance of Identity' in *Identity*, Henry Harris (ed.) (Oxford University Press, Oxford: 1995), and 'Persons, Bodies, and Human Beings', in *Contemporary Debates in Metaphysics*, T. Sider, J. Hawthorne, and D. Zimmerman (eds.) (Basil Blackwell, Oxford: 2008).

²See especially Eric Olson *The Human Animal: Personal Identity without Psychology* (Oxford University Press, Oxford: 1997), 'An Argument for Animalism' in *Personal Identity*, R. Martin and J. Barresi (eds.) (Basil Blackwell, Oxford: 2003), and *What Are We?* (Oxford University Press, Oxford: 2007).

any bodily (or other material) continuities – is dismissed as unwarranted in fact. Having dismissed as unjustified our natural belief in a separately existing subject of experiences, the psychological reductionist concludes that what is important for our survival reduces to the continuation of our mental life, *however* that series of experiences is causally connected, i.e. whether it be continued in one unique body (or mere brain) or through a number of numerically distinct bodies (or brains).

Why does the psychological reductionist reject our natural tendency to believe in a separately existing subject of experiences (a tendency which, he acknowledges, he has himself)? He rejects it on the ground that whereas we have no positive evidence in its favour, we now have much evidence against it. More especially, there is good evidence that all our experiences are (at least) causally dependent for their existence on a proper functioning *material* entity, the brain. From which he concludes that our survival, insofar as it can be justified, cannot be held essentially to consist in a separately existing non-material continuant, despite our natural inclination to believe that it does.

Moreover, as I have already indicated, the psychological reductionist regards the persistence of numerically the same *body* (or other material continuant) as unnecessary for our survival. He claims that, once we have thrown out as unjustified our belief in a separately existing non-material subject of experiences, we should acknowledge that what really matters in our concern to survive reduces to the continuation of our mental life, *whether or not* this mental life is continued in numerically the same body (or mere brain). And in order to back up this claim, the psychological reductionist takes a reliable duplicating process, which he calls ‘teletransportation’. Suppose that an exact printed record is first taken of your body’s (including your brain’s) composition, immediately prior to that body’s total destruction, and that this record is then electronically beamed to Mars, where a replica body is assembled from completely fresh material. Here, says the psychological reductionist, we have a duplication process that wholly destroys bodily, including brain, continuity, without destroying the continuation of mental life or (as he more frequently calls it) *psychological continuity*, viz. that succession of experiences that are linked together by the causal relations that are necessary for the retention of memories, personality traits, and intentions. As the psychological reductionist sees it, since psychological continuity is here preserved, this duplication process as successfully secures what matters in our survival (so far as it can be justified) as does the continuation of our mental life’s *normal* cause, viz. the persistence of numerically the same brain.

For my part, I think this thought experiment does convincingly show that the result of teletransportation is the continuation on Mars of the life that had earlier begun on Earth. Hence, I agree with the psychological reductionist that bodily or mere brain continuity cannot be what really matters for the preservation of that life. And since psychological continuity is not destroyed in the teletransportation process, the psychological reductionist *further* contends that it is this latter continuity, psychological continuity, which is of crucial importance to our survival, so far as it can be justified.

This, very briefly, is how the psychological reductionist reaches his conclusion that what our survival truly consists in – i.e. given that we cannot defend our belief in a separately existing subject of experiences - is psychological continuity, however that continuity is caused.

Now against this view, I maintain that far from the justified content to our belief in survival just consisting in a series of causally connected mental events, the psychological reductionist ought to recognize that the true view of our survival requires us to *distinguish* between the identity of ourselves (as subjects of experience) and the identity of our lives. Only individual lives can even plausibly be thought of as consisting in a series of mental events, suitably causally connected, i.e. in psychological continuity.

My grounds for maintaining that we should distinguish between the identity of selves, or subjects of experience, on the one hand, and the identity of our lives, on the other, stem from this. Although it would nowadays be widely accepted that we are born with a *tabula rasa* with respect to ideas, I submit that we do think of ourselves as coming into the world with a specific *fundamental nature*, in respect of cognitive, conative, and affective capacities (presumably inherited, in some determinate way, from our parents). How our lives develop depends, we hold, on the interaction between two factors. On the one hand, the subject's given fundamental nature and, on the other, the particular sense experiences that this subject has during that life (together with the type of gross body in which that specific fundamental nature is realized).

Suppose that by a reliable process resembling teletransportation – let us call it 'telereproduction' – there is started up, following the natural ending of a person's life on Earth, a wholly *new* life on Mars by means of the same set of instructions as had made possible the original life on Earth. Accordingly, with telereproduction, when the adult body of the subject on Earth has ceased irreparably to function, the duplicating machine is employed to send the electronic record of that body (including its brain) as it was at birth to another machine on Mars, which constructs a replica body of the subject, as it had first existed on Earth. Here, it seems to me, we can legitimately say that, following the ending of the life of a given subject (on Earth), telereproduction has begun *another* life of that same subject (on Mars). Hence, like teletransportation, telereproduction can duplicate a body, including its brain; but, in the case of telereproduction, it duplicates the body as it was at the *start* of a subject's life, i.e. at birth (or some other appropriately early date at which the fundamental nature can be regarded as originally laid down). When it does so, the process can be seen as giving us an example of the same subject living different – and, hence, multiple – lives. In the particular case in hand, one subject can be seen as having two lives: there would first be a life on Earth and, then, a life on Mars¹.

¹For ease of exposition, I consider only cases of *diachronic* duplication (the same subject living multiple lives at *different* times); I do not consider cases of *synchronic* duplication (the same subject living multiple lives at the *same* time). I examine both types of duplication (and in relation to psychological reductionism and my own rival theory) in 'Our Survival', *Res Cogitans* vol. 3 (2006), <http://www.rescogitans.sdu.dk/artikler/index.html>

In speaking here of two lives of the same self or subject, we would mean something along the following lines: that the same fundamental nature (originally encoded in a token human brain) which was responsible, in conjunction with a certain series of sense experiences, for giving a life to a particular self or subject on Earth is now responsible, through its being encoded in another token human brain and in conjunction with a different series of sense experiences, for giving another life to that same self or subject on Mars. Analogously, when we talk of two performances of the same piece of music, e.g. the same piano duet, we mean something like this: that the same set of musical instructions (internalized, via copies of the score, by the performers) which is responsible, in conjunction with a certain interpretation, for producing a performance of the particular duet is also responsible, by its being internalized by another pair of performers (for example), and in conjunction with a different interpretation, for producing another performance of that same duet. An essential condition for enabling us to distinguish between musical works and their performances is the existence of a reliable method for encoding musical sounds in a notational form, viz. in a form that permits multiple performances. Similarly, telereproduction enables us to distinguish between selves or subjects and their lives because it is a reliable process for encoding in notational form a subject's fundamental nature and replicating it in another body. But whereas the distinction between musical works and their performances arguably only arose as a result of the discovery of a reliable method for encoding musical sounds, the distinction between subjects and their experiences – and, thereby, the possibility of distinguishing between subjects and their lives - is one that, as the psychological reductionist admits, is *already* inherent in our thought.

In sum, my argument is this. Since the psychological reductionist accepts that, in seeking to determine what the content of the belief in our survival can justifiably consist in, it is legitimate to imagine the operation of a reliable duplication process *during* a subject's life (as with teletransportation), he should also accept that it is legitimate to engage in a parallel thought experiment at the very *start* of a subject's life (as with telereproduction). When we do so, I have maintained that the relationship between selves or subjects and their lives would be seen as importantly analogous to the relationship that we acknowledge exists between musical works and their performances. We would distinguish between subjects and their lives as we distinguish between musical works and their performances. And having once explicitly recognized the distinction between subjects and their lives, it would be a mistake to regard a subject's identity or survival as in some way *reducible* to a series of causally connected experiences, as the psychological reductionist contends. It would be a mistake because the survival of a self or subject cannot then be said just to consist in those phenomena that go to constitute the experiences of a particular *life*, any more than the survival of a piece of music can be said just to consist in those phenomena that go to constitute the sounds of a particular *performance*. On the contrary, the survival of a self, or subject of experiences, depends on the existence of a record of its fundamental nature which is appropriately

causally related to that subject's conception (so making possible, in conjunction with a set of sense experiences, a life of that subject), just as the survival of a piece of music depends on the existence of its score, or some other record of the music's structure, which is appropriately causally related to that work's creation (so making possible, in conjunction with a set of players, a performance of that piece of music).

'What a reductionist denies is that the subject of experiences is a *separately existing entity*, distinct from a brain and body, and a series of physical and mental events'¹. This is the central reductionist claim. My argument has been that a thought experiment, parallel to the one that the psychological reductionist himself employs in seeking to convince us that the true view of our survival should not be tied to a given brain or body, ought in consistency to convince him that his own central claim is almost entirely the reverse of the truth.

II.

So much for the case against psychological reductionism. What now about animalism, or the biological approach, in the light of the distinction here defended between subjects and their lives?

The animalist brings forward two different considerations – call them A and B - in support of his view that we are animals, and so have the survival conditions suitable to animals:

A) The first consideration itself splits into two halves. On the first half, there is the way that we talk about e.g. our loved ones when they fall into a persistent vegetative state. We talk about them as still alive, and we identify them by their names. We say things like: 'Charles is now in a persistent vegetative state; all hope of his resuming consciousness is gone'. This shows, the animalist contends, that we do not think of a subject as having ceased to exist *merely* because the experiences of its life have irreparably ceased (however much we may then think of that subject's life – reduced now to the life of a human vegetable – as no longer possessing any intrinsic value). On the second half, there is the way that we talk when we are shown X-rays etc. of how we were as an embryo or foetus. We talk of ourselves as already alive, saying things like: 'That's a photo of Mary when she was in the womb; she's certainly grown into a fine young lady'. As the animalist sees it, both these examples show that we do *not* take psychological continuity to be necessary for our continued existence. We do not since, in neither condition, are there *any* conscious states and, hence, none that could be connected to our existence when we do have experiences. On the contrary, claims the animalist, it is our belief that we persist as human animals which explains why we talk in these ways, i.e. why we talk of ourselves as already alive in the womb and as continuing to exist in a persistent vegetative state.

¹Parfit *Reasons and Persons*, p. 223 (italics original).

But these two examples, even if they are successful against psychological reductionism, have no force against the belief in a separately existing self or subject of experiences of the type I have defended. Take the case of Charles who falls into a persistent vegetative state. It is consistent to hold *both* that the life of that human being has lost all its intrinsic value when the possibility of any mentality has ended – has, to all intents and purposes, ceased – *and* that the subject of experiences can neither have ceased thereby to exist nor be identified with the continuously existing human animal. While the animalist can rightly affirm that the persistence of a human animal is dependent on the continued functioning of a biological organism and not on psychological continuity, this insistence is compatible with holding that the subject of experiences is a separately existing entity whose continued existence is *not* determined by the persistence conditions of the human animal. No doubt, a human being's life loses its intrinsic value with the irreparable cessation of all its mentality (if, indeed, the human being, rather than the human animal, can still be thought of as alive with the termination of all its possible experiences); but the subject of those experiences has not ceased to exist, nor will it when the given human animal ceases to exist with the termination of biological functioning.

A similar reply goes for the case where, during a life, we think of ourselves as having grown into a healthy adult from an embryo or foetus. Although it does indeed seem plausible to claim that we already think of a given human being as alive at a relatively early stage of foetal development, despite there being no psychological continuity between this stage and later stages (there are no signs of consciousness in an embryo or ten-week old foetus), it is not the case that the difficulties that this raises for the psychological reductionist have force against the distinction, here drawn, between the self or subject and its possible lives. For in supposing ourselves to have started up a life at a relatively early stage of foetal development, we are taking it that our fundamental nature is already to be found there encoded, albeit in – as we might say – an embryonic state. Accordingly, there is no inconsistency in maintaining *both* that we have started up a life as an embryo or at a relatively early foetal stage *and* that the continued existence of the self or subject of experiences is not to be identified with the persistence of any given human being or animal.

In sum, whether we consider the human animal at the start of its existence or near its end (and after mentality has ceased), any difficulties that may arise for the psychological reductionist do not arise for the alternative account being offered here.

B) The second consideration that the animalist brings forward in support of his own view is often referred to as 'The Thinking-Animal Argument'. Say you claim that you are not identical with any human animal. In which case, says the animalist, it follows that when you are sitting in a chair and thinking, there will *also* be a human animal sitting in that chair. Now since it would be unrealistic to deny that this human animal is thinking (it has an operating cerebral cortex), it follows that you will have to hold that there are *two* beings sitting in that chair, yourself and the human animal, both of which are having qualitatively

the same thoughts. This is not credible. The only reasonable thing to say is that you are identical with the animal that is now sitting in the chair. But this is to concede that animalism is the correct account of what we are.

However, on the theory that I am defending, it is not literally the case *both* that there is a human animal sitting in that chair *and* that the self or subject that you are is sitting in that chair. Certainly, there is a human animal that is sitting and thinking in that chair; for there is a human animal that is sitting in that chair, and that animal is now having conscious states. Moreover, these conscious states do constitute part of a life that *you* are currently leading as a particular human being. You have a life (anyway as things now are) only by means of the germination and growth of a human animal. But it is at best misleading to say that the self or subject that you are is now sitting in that chair, as it would be at best misleading to say that Beethoven's Fifth Symphony is now in the Royal Albert Hall. While a performance of a piece of music can be spatially located, the piece of music itself is not a spatial object. Similarly, while the life that you are now leading can be spatially located (at a given time and through time), the self or subject that you are cannot be spatially located. Hence, it is at best misleading to say that you are now thinking in that chair. It is at best misleading, if it is merely taken as a shorthand way of saying that there is a human being that is currently realizing a life of yours, and it is now thinking in that chair. It is plain wrong if it is taken as saying that the self, whose life is now being lead by the human being in that chair, is itself sitting and thinking in that chair. Again, the musical comparison helps to clarify the point. To say that Beethoven's Fifth Symphony is now playing in the Royal Albert Hall is at best misleading, if it is merely taken as a shorthand way of saying (truly) that there is an orchestra that is now performing that symphony and it is doing so in the Royal Albert Hall. But it is plain wrong if it is taken as saying that the symphony that is now being performed by a given orchestra in the Royal Albert Hall is itself playing in the Royal Albert Hall.

My conclusion is that none of the arguments that the animalist brings forward against the psychological reductionist, and in favour of his own view, are effective against the alternative view that I am supporting. Given the way that we think about ourselves, I maintain that the most defensible view of our survival is that it resides neither in a causally connected series of experiences (as the psychological reductionist affirms) nor in the persistence of a human biological organism (as the animalist affirms), but in a significant *further* fact, viz. in the persistence of a self or subject that is genuinely separate from both.

III.

I end by sketching an important consequence of this distinction between the self or subject of experiences and its possible lives. I claim that by making the distinction, we can conceive of ourselves as possessing the capacity for

exercising a degree of self-determination that goes significantly beyond what is presently allowed for in compatibilist accounts of free will.

However, it might now be objected that on my account of the self or subject of experiences, this self can neither think, nor decide, nor act. If the self is identified as a continuant that is capable of multiple lives (analogous to the way in which a piece of music is capable of multiple performances), then it surely follows that this self cannot *do* anything (any more than a piece of music can *make* sounds). How, therefore, can I claim that this subject, this I, can possess *any* degree of self-determination? The answer is that we need to distinguish between, on the one hand, the self as a continuant with a specific fundamental nature and, thereby, with the potentiality for having experiences, and, on the other, the actual realization of this self *as a particular human being*. It is the self, as it is realized in a given life, i.e. as a particular human being, that thinks, and decides, and acts. And my contention is that, along with the possibility of multiple realizations, the self, within a given life, will correspondingly possess a greatly increased capacity for self-determination. Just as, on the account that I have been defending, the subject's existence transcends any particular life, so the freedom that the subject is capable of exercising, *in* a life, can transcend its ability to act unconstrained on the basis of an understanding of its fundamental nature as this happens to manifest itself solely in the exigencies of this one particular life. For, through a careful study of its earlier lives, the subject will be provided with a decisive opportunity for directing its present life in more fulfilling ways than could be expected from any study of its present life alone.

This is so because the subject, by exercising its judgment on how it has behaved in its past lives, can now develop a grasp of its own fundamental nature which is far richer and deeper than anything that could reasonably result from an observation of its particular manifestations in this individual life alone. Consequently, the agent's choices, within its present life, can be based on a profound knowledge of its own true nature. An analogy with musical works is again helpful. By comparing and contrasting different interpretations of a given work, one can come to a much richer and deeper appreciation of the music itself than by listening to any single interpretation. In fact, it would seem that for us the *only* way to grasp the heart of a piece of music is to hear various interpretations played: there is no more direct route to grasping its fundamental nature. Similarly, it is by the subject, within its present life, examining the various ways that its fundamental nature has been manifested, not only in this life but in the multifarious circumstances of its different past lives, that can alone enable the subject to develop a penetrating appreciation of its true nature.

By thus gaining a deep and wide-ranging understanding of its fundamental nature, the subject is enabled to stand outside those decisions that would otherwise result from the subject's interaction with the circumstances obtaining in its present life. Instead, the subject will be enabled to reach its decisions on the basis of a substantially increased knowledge of its fundamental nature; and so its actions can manifest a greatly enhanced recognition of what, in the circumstances of its present life, is most likely to accord with that nature. Certainly, on this account, the subject, even when it acts on something

approaching a full knowledge of its own true nature, will still be acting on the basis of a naturally produced decision, and not through a temporally uncaused decision. Nonetheless, this account does distinguish between those decisions that simply arise from the agent's consideration of itself and its circumstances in this life alone, and those decisions that are based on a profound appreciation of its fundamental nature (an appreciation that is made possible from studying how that nature has manifested itself in the variety of its different lives to date). To be fully free, on this account, would be to act unconstrained on the knowledge of one's fundamental nature as it really is, and not on the basis of a knowledge of that nature as it merely appears to be from considering those desires and decisions that are generated by means of the circumstances obtaining in one's present life alone.

The upshot is that decisions that simply result from reflecting on what has happened to the subject in its present life can, at best, only amount to the ensuing actions being based on the subject's understanding of its fundamental nature as this now appears to exist. But when the subject, within a given life, acts on decisions that have arisen from reflecting on its nature as this has been manifested in the various circumstances of its different lives to date, then its ensuing actions can be based on a greatly enriched appreciation of its nature as it really is. In this latter case, the subject can rightly be said to be closer to producing its actions entirely *of itself*, that is, on the basis of decisions that arise from a profound grasp of its own true nature, and not on decisions that merely reflect an understanding of how its fundamental nature happens to have manifested itself within the contingencies of its present temporal existence alone. When the subject acts unconstrained on the basis of a deep and wide-ranging appreciation of its nature as it really is, it can be said to be approaching full autonomy; and the deeper and wider this appreciation extends, the nearer the subject comes to complete *self*-determination.

Here is an overall summary of my argument. By articulating the distinction between the self or subject and its series of experiences – a distinction that is already inherent in our conception of ourselves – we can show that neither psychological reductionism nor animalism provides an adequate account of our survival. Instead, we can defend the belief that we are separately existing selves or subjects of experience. And, by so doing, we can also show that this self, within a given life, is capable of a freedom that transcends any of the desires and decisions that would be generated solely from the self's consideration of the circumstances obtaining in that life. The self, rather, has the capacity to motivate itself not simply on the basis of an understanding of its fundamental nature as this merely *appears* to itself from the circumstances within its present life alone, but on the basis of a deep and extensive appreciation of that nature as it has manifested itself through all the various situations in its past lives. In this way, each of us can act on decisions that reflect a genuine recognition of our fundamental nature as it *really* is; and so, when not constrained by outside forces, we can truly be said to approach full self-determination.

CHAPTER EIGHTEEN

Ramsey and Laconicism

Kiriake Xerohemona

...we do not hope to learn an infallible means of distinguishing truth from falsehood but simply to know what the word 'true' means. It is a word which we all understand, but if we try to explain it, we can easily get involved, as the history of philosophy shows, in a maze of confusion. - F. P. Ramsey.¹

Deflationism is the most general and common term used to characterize a whole family of theories about truth. Deflationism has been associated with many positions in contemporary discussions of truth, and has been attributed retrospectively to Aristotle, Frege, Ramsey, Quine, and even to Tarski. I argue here that Ramsey is clearly committed to the core deflationist theses. At the same time, I dispel two perceptions: that Ramsey is nothing but a redundancy theorist, and that he is really a correspondence theorist in disguise. (The other retroactive attributions, I believe, are mistaken; but I cannot argue them here.) The goal of the first part of this paper is to articulate the core theses of deflationism by drawing on the work of contemporary deflationists; the goal of the second part, to show that Ramsey satisfies the three theses and is clearly a 'laconicist' regarding truth since his account is as the word suggests, lean, concise and elegant. Deflationism is often contrasted with 'substantive' or 'inflationary' approaches taken to include correspondence, coherence, and pragmatic theories. Substantive theories of truth presuppose that truth is a substantive concept: that truth is a real, robust, metaphysically interesting property, and 'is true' a real predicate in the sense that it attributes a genuine property to statements, beliefs, or propositions; that the concept of truth can be made explicit; that inquiry into the nature of truth is a meaningful and useful enterprise that will (eventually) provide answers to perennial philosophical problems that make use of the concept of truth. It is not easy to offer a simple and straightforward explanation of what deflationism is, in part because the term 'deflationism' is not used consistently, and has been given to older accounts of truth retrospectively. There are many variants of contemporary deflationism, most coming with their own labels: Horwich's 'minimalism', Soames's version, Field's 'pure disquotational truth', Ramsey's 'nihilism', or 'redundancy theory', Grover et al.' 'prosentential theory', etc.² Taking 'deflationism' as the umbrella term, I will try to explicate the theory drawing mainly on three influential contemporary proponents of deflationism: Horwich, Soames, and Field.

¹Ramsey in (Rescher & Majer, 1990), p. 6.

²(Horwich, 1998); (Soames, 1999); (Quine, 1970); (Quine, 1990); (Field, 1994); (Ramsey, 1927); and Ramsey in (Rescher & Majer, 1990); (Grover, Camp & Belnap, 1975)

Perspectives of Contemporary Deflationists

Horwich is among the most prominent figures in the deflationist discussion. He advocates a version of deflationism which he calls 'minimalism'. Horwich holds that 'true' is a predicate and that truth is a property; it is just that truth is not a naturalistic property; that is, a property 'whose nature can be specified by a naturalistic reduction or by laws relating it to natural phenomena'.¹ Horwich takes propositions as the primary bearers of truth. Here are some of Horwich's claims:

- H1: *'traditional investigations into the underlying nature of truth have been misconceived' for it is false that 'truth has some hidden nature awaiting our discovery'*.²
- H2: *'... the traditional attempt to discern the essence of truth - to analyze that special quality which all truths supposedly have in common - is just a pseudo-problem based on syntactic overgeneralization. Unlike most other properties being true is unsusceptible to conceptual or scientific analysis. No wonder that its "underlying nature" has so stubbornly resisted philosophical elaboration; for there is simply no such thing'*.³
- H3: *'Truth is not a normal property and traditional investigations into its underlying nature have been misconceived'*.⁴
- H4: *'The utility and function of the truth predicate is not what of most predicates'... 'The role of truth is not what it seems. In fact, the truth predicate exists solely for the sake of a certain logical need'*.⁵
- H5: *'The proposition that p is true if and only if p' will provide 'the basis for accounts of the meaning and function of the truth predicate, of our understanding it, of our grasp upon the concept of truth, and of the character of truth itself'*.⁶

Soames is another contemporary defender of deflationism who also accepts propositions as the bearers of truth. Here are some of his dicta on deflationism:

- S1: *it is 'the view that truth is a philosophically uncontentious concept, the content of which is grasped by recognizing the equivalence of a putative truth bearer with the claim that it is true'*.⁷
- S2: *'The leading idea behind deflationism about truth-namely, that claims of the sort "It is that that S" and "The proposition that S*

¹(Horwich, Summer 2002)

²(Horwich, 1998), p. 6.

³(Horwich, 1998), p. 5.

⁴(Horwich, 1998), p. 15.

⁵(Horwich, 1998), p. 2.

⁶(Horwich, 1998), p. 37.

⁷(Soames, 1999), p. 229.

*is true” are trivially equivalent to S and that this equivalence is in some sense definitional of the notion of truth’.*¹

S3: *‘sweeping, philosophically contentious doctrines about reality and our ability to know that it cannot be established by analyzing the notion of truth’.*²

S4: *‘the content of the claim that a putative truth bearer is true is equivalent to the truth bearer itself, a fact that endows the truth predicate with an important practical and theoretical utility’.*³

Field, the proponent of the ‘pure disquotational truth’ version of deflationism, takes utterances and ‘states of thinking’ (*sic*) to be the truth bearers. Acknowledging Quine as having introduced the phrase ‘diquotationalist’, Field states that:

F1: *‘We need a notion of truth for purely logical reasons: “true” is simply a device for infinite conjunction (or disjunction), a device which would be of little use were we to have in our languages other devices of infinite conjunction’.*⁴

F2: *‘The only useful notions of truth are disquotational truth and various other notions definable from it using rather limited additional resources’. ‘disquotational truth’ has two characteristics: [a] it is defined only for sentences that one understands, and [b] the property of those sentences which it defines is one that a sentence has or fails to have independently of the way that the sentence is used by speakers’.*⁵

F3: *‘“For disquotational truth” there is an “intimate connection” between saying of one’s utterance of ‘snow is white’ that is true and saying that snow is white: it is identity’.*⁶ *‘a person can meaningfully apply “true” in the pure disquotational sense only to utterances that he has some understanding of; and for such an utterance u, the claim that us is true (true-as-he-understands-it) is cognitively equivalent (for the person) to u itself (as he understands it)’.*⁷

F4: *‘...purely disquotational truth is a purely logical device, and its nature is exhausted by the totality of T-sentences’.*⁸

¹(Soames, 1999), p. 231.

²(Soames, 1999), p. 231.

³(Soames, 1999), p. 231.

⁴Field in (Macdonald & Wright, 1986), p. 57. Also in (Field, 1994), p. 11.

⁵Field in (Macdonald & Wright, 1986), pp: 59, 58.

⁶Field in (Macdonald & Wright, 1986), p. 71.

⁷(Field, 1994), p. 2.

⁸(Field, 1994), p. 2.

Contemporary Deflationism Defined

Putting aside the differences among these deflationists's views, three ideas appear to be held by all three:

1. **The Alethic Thesis:** The Alethic (or equivalence) thesis states what it means for a truth bearer to be true, it articulates the core meaning of 'true'. There are variants of the Alethic thesis, depending on what the truth bearer is taken to be, and how the relation between stating the truth bearer and stating that the truth bearer is true is interpreted. Some deflationists accept sentences as truth bearers, some accept utterances, while others accept propositions. Moreover, some deflationists maintain only logical equivalence between the proposition that a truth bearer is true and the truth bearer itself, while others claim identity of meaning.

Here are some variants of the thesis: To assert that a statement is true is just to assert the statement itself; or stating that a proposition is true is the same (means the same) as stating the proposition; or 'It is true that p ', 'the proposition that p is true' are equivalent to ' p '; the sentence ' p '; the sentence ' p ' is true iff p ; or a sentence that states that a sentence is true is equivalent to the sentence itself. (H5, S2, F3)

2. **The Positive Thesis:** The Positive Thesis refers to the function and utility of the truth predicate. The main functions of the truth predicate are two: logical and practical. The truth predicate adds to the expressive power of natural language since it is mainly used as a device for expressing generalizations and for making infinite conjunctions and disjunctions. We need the truth predicate to say, for instance, that 'all axioms of arithmetic are true', or that every sentence of the form 'if p , then q , and p , then q ' is true.

In addition, we need 'true' for practical reasons. The truth predicate is needed when we want to express agreement with someone without repeating what the person said, as for instance in 'what you said about the article is true', or for semantic ascent. The truth predicate enables us to move from talk about language to talk about reality. For example, we can move from 'the sky is blue' is true to 'the sky is blue', since by attributing truth to the sentence, we attribute blueness to the sky. (H4, S4, F1)

3. **The Negative Thesis:** The Negative thesis has to do with the nature of truth and the concept of truth. It is to be noted that the Negative Thesis is formulated vaguely by modern deflationists. There are several different versions:

Truth does not have a metaphysical nature. Truth is not a property, or a substantive property. Truth does not have an 'underlying nature', 'a hidden structure waiting our discovery'. Analysis of the nature of truth is a misguided enterprise; the nature of truth does not play a significant role in explaining or elucidating such concepts as reference or meaning. (H1, H2, H3, S1, S3, F4).

Ramsey as Precursor of Deflationism

The true ancestor of contemporary deflationism is Ramsey whose account of truth clearly satisfies the three theses. Ramsey, in the 1920s, originated the key idea, and offered a detailed explanation and justification for the theory. Ramsey acknowledges that he found a similar idea in the works of Aristotle and Kant.¹ Ramsey, applies 'true' and 'false' to mental states that have assertive character and carry propositional content or have what he calls 'propositional reference'.² Ramsey's position on the three theses may not appear to be quite straightforward; but this is because of two common misinterpretations of Ramsey's theory: an overinterpretation, which makes him say too much with respect to truth, and an underinterpretation, which makes him say too little. The overinterpretation is that Ramsey is a correspondence-theorist. The underinterpretation is that Ramsey is a redundancy theorist. Both, however, are misinterpretations - as we shall see as my discussion of Ramsey's attitudes on the three deflationists theses proceeds.

Ramsey and the Alethic Thesis

Ramsey formulates his theory of truth in the manuscript *On Truth* (1927-1929) and in the essay 'Facts and Propositions' (1927).³ The basic idea behind Ramsey's theory of truth is:

- [R1] *The most certain thing about truth is that 'p is true' and 'p', if not identical, are equivalent.*⁴
- [R2] *We can say that a belief is true if it is a belief that p, and p.*⁵
- [R3] *A belief that p is true if and only if p.*⁶
- [R4] *Truth, we say, is when a man believes that A is B and A is B, whether or not such an occurrence can be accurately described*

¹Ramsey in (Sahlin, 1990), p. 236, ft. 4.

²He defines propositional reference in terms of a belief. 'A belief is necessarily a belief that something or other is so-and-so'. The expression 'the belief that', or 'that is so-and-so' conveys the characteristic of propositional reference. In 'I believe that it is raining', the expression 'that it is raining' is the propositional reference of my belief; it is virtue of its propositional reference that a belief has a truth value. Ramsey in (Rescher & Majer, 1990), pp: 7-9. Unfortunately, Ramsey did not produce a fully developed account of propositional reference.

³Ramsey in (Ramsey, 1927).

⁴Ramsey in (Rescher & Majer, 1990), p. xviii (editor's introduction).

⁵ibid, p. 9.

⁶ibid, p. 13.

*as a correspondence between two facts; failure to describe it in terms of correspondence cannot show that it never occurs and is not what we mean by truth.*¹

In 'Facts and Propositions' in the course of an analysis of the notion of judgment, Ramsey writes:

[R5] *...before we proceed further with the analysis of judgment, it is necessary to say something about truth and falsehood, in order to show that there is really no separate problem of truth but merely a linguistic muddle. Truth and falsity are ascribed primarily to propositions. The proposition to which they are ascribed may be either explicitly given or described. Suppose first that it is explicitly given; then it is evident that 'It is true that Caesar was murdered' means no more than that Caesar was murdered, and 'It is false that Caesar was not murdered' means that Caesar was not murdered. They are phrases which we sometimes use for emphasis or for stylistic reasons, or to indicate the position occupied by the statement in our argument.*²

It is clear from the above that Ramsey's dicta point to the Alethic Thesis. However, Ramsey's commitment to the thesis is disguised a bit by a common over-interpretation. After Ramsey articulates his theory that 'it is truth that p', or 'the proposition that p is true', or 'it is a fact that p' is just another way of saying that p, he says that his theory will no doubt be called a correspondence theory: 'Although we have not yet used the word "correspondence" ours will probably be called a Correspondence Theory of Truth'.³ And he continues, 'we cannot describe the nature of this correspondence until we know the analysis of propositional reference, of 'believing that A is B'. Only when we know the structure of belief can we say what type of correspondence it is that unites true beliefs with facts'.⁴ This certainly seems to point to correspondence. But Ramsey does not stop there; he immediately starts criticizing the idea of correspondence. First of all, he argues that it is 'an inaccurate popular explanation' to construe truth in terms of correspondence. Second, he specifically states that it is doubtful that there will be 'any simple relation of correspondence applicable to all cases'.⁵ Ramsey believes that his is 'a clear definition of truth which escapes all these difficulties' precisely by avoiding appealing to a notion of correspondence at all.⁶ The view that a belief that p is true if and only if p that does not appeal to any notion of correspondence may

¹ibid, p. 12.

²Ramsey in (Mellor, 1990), p. 38.

³Ramsey in (Rescher & Majer, 1990), p. 11.

⁴ibid, pp: 11, 13-14.

⁵ibid, p. 12.

⁶ibid, p. 11.

look like an 'obvious truism,' 'so obvious that one is ashamed to insist on it,' 'a trivial formalism' but it is much clearer, and avoids the pitfalls of a relation of correspondence-of-a-belief-with-a-fact.¹ As Ramsey writes: 'Talk of correspondence, though legitimate and convenient for some purposes, gives, in my opinion, not an analysis of truth but a cumbrous periphrasis, which it is misleading to take for analysis. To believe truly is to believe that p when p, and there is no need to recast this definition in terms of correspondence, unless indeed, some sort of correspondence is essentially involved in the notion of 'believing that p,' a question which we must leave till we come to the analysis of propositional reference'.²As we can see from the above that Ramsey is not a whole-hearted correspondence theorist. It is also clear that he strongly supports the Alethic Thesis.

Ramsey and the Positive Thesis

Ramsey's commitment to the Positive Thesis is disguised a bit by an even commoner under-interpretation, according to which Ramsey's is a 'redundancy' theory. I believe that this interpretation is also mistaken. The issue is whether Ramsey thought the truth predicate was redundant. I believe there is evidence that he did not. Ramsey's main preoccupation in *On Truth* is to 'elucidate the terms true and false' as applied to beliefs, and judgments, and to explain the meaning of the word 'true'. To this effect he proposes his view using directly or explicitly asserted propositions, such as, 'It is truth that the earth is round,' or 'It is true that Caesar was murdered.'³ In such contexts, it is enough to remove 'it is true that' without any semantic loss. 'It is true that the earth is round' and 'the earth is round' are equivalent. However, he is well aware that these are not the only circumstances in which we use the term 'true', and we need it in those other circumstances because natural language is ill-equipped to handle without the use of 'true'. More specifically, Ramsey considers examples of indirect-truth ascription contexts in both 'Facts and Propositions'⁴ and in 'On Truth' where he states that

... As we claim to have defined truth we ought to be able to substitute our definition for the word 'true' whenever it occurs. But this difficulty⁵ we have mentioned renders this impossible in ordinary language which treats what should really be called pro-sentences as if they were pronouns. The only pro-sentences admitted by ordinary language are 'yes' and 'no', which are regarded as by themselves expressing a complete sense, whereas 'that' and 'what' even when functioning as short for sentences always required to be supplied with a verb: this verb is often 'is

¹Ramsey in (Rescher & Majer), pp: 14, 19-20, 22.

²ibid, p. 19.

³Ramsey uses the latest example in 'Facts and Propositions'.

⁴Ramsey in (Ramsey, 1927), p. 10.

⁵The 'difficulty' here refers to replacing the truth predicate in indirect speech contexts.

true' and this peculiarity of language gives rise to artificial problems as to the nature of truth, which disappear at once when they are expressed in logical symbolism, in which we can render 'what he believed is true' by 'if p was what he believed, p'.¹

Ramsey resorts to a semi-formal language that uses variables, and quantifiers to represent indirect truth ascriptions, and eliminate the truth predicate. Let us consider such an ascription:

[B]: What he believed is true. (Following Ramsey's paraphrase) [B] becomes:

[BF]: For all p, if he believes p, then p.

Such a swift move is promising but not unproblematic. For [BF] appears to be ill-formed (the last occurrence of 'p' in [BF] is a variable, and variables are grammatically singular terms), so 'is true' must be added to restore its proper grammatical structure. With the reintroduction of the alethic predicate [BF] becomes [BFT]

[BFT]: For all p, if he believes p, then p is true. (which leads to the objection that the truth predicate has not been eliminated.)

Ramsey acknowledges the appearance of circularity in [BFT] and suggests to take the last occurrence of 'p' in [BF] as a sentential or propositional variable-as such it already contains the verb, and there is no need to add 'is true'. He, also, claims that sentential variables along with the addition of the alethic predicate can be considered 'pro-sentences' which function as pronouns function for nouns. However, Ramsey's swift response regarding the eliminability of the alethic predicate in a semi-formal language leaves some questions open. For it is not clear that the alethic predicate has been eliminated by such a paraphrase. To show this, Ramsey must offer an account of quantifying over propositions, an account of propositional variables that do not themselves depend on the concept of truth, and an account of the character and function of pro-sentences.² What is clear from the above is that Ramsey does not really claim that the 'true' term is redundant in all contexts in ordinary-language, as he is often portrayed as having said. 'true' is eliminable in direct truth ascriptions; it is not eliminable in indirect truth ascriptions unless it is translated in a quasi-formal language, but as we saw, even then it is not entirely clear that it can be eliminated altogether. It would be understandable, and even excusable, to label Ramsey a redundancy theorist, if the only available evidence were the 1927 'Facts and Propositions'. But given the above, it is

¹Ramsey in (Rescher & Majer, 1990), p. 10.

²(Williams, 1976), (Grover et al, 1975) are motivated by Ramsey's theory. They focus on the analysis of indirect truth ascriptions and offer accounts of the nature and function of pro-sentences. For the prosententialists the alethic predicate is a useful logical tool for the construction of complex pro-sentences.

quite surprising and unfortunate that this misinterpretation continues even after the publication of Ramsey's work *On Truth*. The editors of this work not only keep using the misleading term 'redundancy theory' to describe Ramsey's project and so perpetuate an inaccurate reading of Ramsey's theory,¹ but also say that 'Ramsey was the first to mention the redundancy thesis to the effect that 'is true' is seen as invariably eliminable in all contexts of its application to propositional objects since any such application is simply tantamount to asserting the proposition itself'.² Ramsey clearly accepts the Positive Thesis. He agrees that the truth predicate adds to the expressive power of natural languages. He agrees that it allows us to express indirect ascriptions in a simple and straightforward way and therefore, its role is invaluable in natural language. It is safe to conclude that he would accept the Positive Thesis.

Ramsey and the Negative Thesis

The Negative Thesis is formulated very vaguely by contemporary deflationists, and any doubt whether Ramsey accepts the Negative Thesis is due to this vagueness. There are several plausible interpretations: that truth has no metaphysical nature, that there is no genuine property of truth, that there is no need for further investigation into the nature of truth, that the nature of truth does not play a significant role in explaining or elucidating such concepts as reference and meaning, and so on. Ramsey agrees that no metaphysical 'nature' of truth needs to be explored. After all, when he criticizes the correspondence theory, he questions the existence of facts, disjunctive facts, and correspondence. Also, he criticizes attempts to define true beliefs in terms of being 'well-grounded' or 'comprehensive'. Ramsey is not interested in the metaphysical nature of truth; but of course, that does not mean that he takes the concept of truth to be superfluous or unimportant. Ramsey takes the essential problem to be not the notion of truth but the notions of beliefs, judgments, and assertions. More specifically, he argues that there is 'no separate problem of truth', 'but merely a linguistic muddle'.³ 'The problem', Ramsey says, 'is not as to the nature of truth and falsehood, but as to the nature of judgment or assertion...It is, perhaps, also immediately obvious that if we have analyzed judgment we have solved the problem of truth'.⁴ This might be the reason why in 'Facts and Propositions', Ramsey presents his account of truth in such a condensed way, while talking about the proper analysis of the nature of

¹For instance, Soames in (Soames, 1999) states that: 'according to the classical redundancy theory (which he attributes to Ramsey, Ayer, Prior, Grover, and Belnap), there is no such property as truth, and the predicate *is true* is not used to describe anything' (p. 232), hence the label 'redundancy theory of truth'. Brendel in (Chapuis & Gupta, 2000), p. 40 says about Ramsey's theory: 'in redundancy theories of truth "is true" is considered superfluous, useless operator. To say of a sentence that is true is to say the same as just stating the sentence. The notion of truth is therefore redundant and without any loss of expressive power one can do without it'.

²Ramsey in (Rescher & Majer, 1990), p. xv, (editor's introduction).

³Ramsey in (Ramsey, 1927), p. 38.

⁴Ramsey in (Ramsey, 1927), p. 39.

judgments and assertions. There is additional evidence in *On Truth* where Ramsey claims that ‘truth depends on and is defined in terms of ‘propositional reference’; ‘Truth’, Ramsey says, ‘is when a man believes that A is B and A is B’. So, we need to know what it means to believe that A is B. He adds that ‘it is essential to realize that truth and propositional reference are not independent notions requiring separate analysis, and that it is truth that depends on and must be defined via reference not reference via truth’;¹ and that ‘we shall clearly not have finished with truth until we have got to the bottom of propositional reference’.² Now, these suggest that Ramsey view the problem of truth as secondary to the problem of analysis of the contents of beliefs, judgments and assertions, but this does not suggest that the truth is redundant or vacuous. Ramsey was interested in providing a proper understanding of the meaning of ‘true’. He wrote *On Truth* between 1927 and 1929, when the logical positivist current was skeptical of the notion of truth. This period was not a ‘propitious time for philosophical treatment of the subject’, and ‘elucidation of truth was not a priority’.³ Yet Ramsey not only understood the significance of providing an appropriate definition of a true statement but devoted a considerable amount of time attempting to capture the core meaning of ‘true’. His account is ‘merely a truism’, a ‘platitude’, a ‘trivial formalism that we cannot contradict without absurdity’, and is ‘so obvious that one is ashamed to insist on it’.⁴ Ramsey believes that his truism is the heart and core of all plausible theories of truth and he insists that any investigation into the concept of truth ‘must square with the obvious truism’, i.e., that a belief that p is true if and only if p and must answer what it is to believe that p . Ramsey is not only the clearest anticipator of contemporary deflationism, but in some ways subtler than those who subscribe to deflationism today. His theory seems incomplete since he has not provided a deeper explanation of propositional quantification, propositional variables that paraphrase indirect truth ascriptions. However, his theory does not require substantive ontological commitments and does not rely on an ad hoc stratification of languages which may be thought masterfully to solve formal problems, but which disregards characteristics of the ordinary notion of truth. He does not want to offer a criterion of truth but attempts to capture the core meaning of ‘true’ in its various applications. He talks about the ordinary concept of truth as it applies to beliefs, judgments, and assertions. What is chiefly important about Ramsey’s account is that it satisfies the three theses of contemporary deflationism, and is so much more basic than modern deflationist theories. Since the term ‘deflationism’ is so over-used and poorly defined - I suggest a novel term for Ramsey’s proto-deflationism: laconicism. For his account is indeed, as the word suggests, lean, terse, concise and elegant.

¹Ramsey in (Rescher & Majer, 1990), p. 14.

²ibid, p. 22.

³ibid, p. xiv, (editor’s introduction).

⁴ibid, pp: 12-14.

References

- Chapuis, A., & Gupata, A. (2000). *Circularity, Definition, and Truth*. Indian Council of Philosophical Research, New Delhi.
- Field, H. (1994). 'Deflationist Views of Meaning and Content.' *Mind*, 103.
- Grover, D., Camp, J., & Belnap, N. (1975). 'A Prosentential Theory of Truth.' *Philosophical Studies*, 27, 73-125.
- Horwich, P. (1998). *Truth* (2nd ed). Oxford University Press.
- Horwich, P. (Summer 2002). 'Varieties of Deflationism.' Paper presented at the 14th Annual European Congress of Analytic Philosophy.
- Macdonald, G. & Wright, G. (1986). *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth, and Logic*. Blackwell.
- Mellor, D. H. (ed.) (1990). *Philosophical Papers of F. P. Ramsey*. Cambridge University Press.
- Quine, W. V. (1970). *Philosophy of Logic*. Englewood Cliffs.
- Quine, W. V. (1990). *Pursuit of Truth*. Harvard University Press.
- Ramsey, F. P. (1927). 'Facts and Propositions.' *Proceedings of the Aristotelian Society - Supplementary Volume*, 7, 153-170. (All page references are to Mellor, D. H. (1990).)
- Rescher, N. , & Majer, U. (eds)(1990). *Frank Plumpton Ramsey On Truth: Original Manuscript Materials (1927-1929)*. Kluwer Academic Publishers.
- Sahlin, N. E. (1990). *The Philosophy of F. P. Ramsey*. Cambridge University Press.
- Soames, S. (1999). *Understanding Truth*. Oxford University Press.
- Williams, C. J. F. (1976). *What is Truth?*. Cambridge University Press.

Part 3

Ethics and Political Philosophy

CHAPTER NINETEEN

Sartre and Judith Butler: A Problem of Agency That Existentialism Can't Solve

Katherine Cooklin

Several philosophers have argued that there are deep convergences between Foucault and Butler on the one hand and Sartre on the other (Boileau, 2000; Schrift, 2001). Alan Schrift argues that those inspired by Foucault, like Judith Butler, would be well advised to abandon the exhausted poststructural trajectory and begin looking in Sartre's existential tool box for resources to further their politics. Butler describes her project as remaining within a Foucauldian framework, and Foucault's work could be read as having existentialist overtones in the same way that Schrift reads Butler. Therefore, if Sartre's work could serve as a corrective for Butler, then it could serve as a corrective for Foucault. I will argue, however, that neither can be served by Sartre because they both jettison Sartre's theory of consciousness, the theoretical tool necessary to Sartre's existentialism.

Schrift points out that the legacy of Sartre's work has influenced many French intellectuals, including Foucault, although this influence remains for the most part unacknowledged. 'But surely it should strike us as strange that so many philosophers who pride themselves on their historical sense and sensitivity to historical context should so easily ignore the thirty-odd year dominance of Sartre and existentialism over the French philosophical scene and ignore as well the historical legacy of existentialism as it influenced several generations of French intellectuals. I am thinking here primarily of the French philosophers themselves, philosophers like Derrida, Foucault, Irigaray, or Lyotard, who must have had to work very hard at forgetting that Sartre, or Beauvoir or Merleau-Ponty for that matter, ever wrote philosophy' (Schrift, 2001, 14). Schrift is right in acknowledging Sartre's contribution to contemporary French philosophy. It was Sartre who proclaimed that the empirical ego was not the seat of consciousness, but an object out in the world, better known perhaps by others than by oneself. Sartre also purged consciousness of any transcendental ego. There can be no doubt that Sartre's rejection of the transcendental subject was in large part responsible for the shift in Continental philosophy that celebrates the *Death of Man*.

According to Schrift, Butler's failure to adequately acknowledge Sartre's philosophy is to her detriment. Pursuing the existentialist dimension could provide a more fruitful strategy for avoiding a problem that has plagued her as well as many of the poststructural philosophers, namely, developing an account of agency or the subject that facilitates rather than hinders the adoption of a leftist political agenda (Schrift, 2001, 15). Schrift sees a solution to this

problem in the work of Sartre. Schrift points out that Sartre saw an ‘account of the self as a free project of existence as a central thesis in the development of [his] leftist politics. The subsequent generations of French philosophers have not been as successful at bringing together their philosophical and political positions and for this reason, one must wonder whether, if one was forced to choose allies in a political struggle, one might...prefer to cast one’s lot with the likes of Sartre...than with Althusser [or] Lacan (ibid).

Given Butler’s earlier engagement with Sartre in her *Subjects of Desire*, Schrift wonders why she does not realize the advantage of utilizing the works of Sartre over the works of the poststructuralists like Lacan or Althusser. What Schrift does not seem to realize is the fact that one may prefer to “cast one’s lot” with Sartre instead of Althusser or Lacan simply highlights the incompatibility of these particular poststructural theories with emancipatory politics. It does not show that Butler could cast her lot with Sartre while remaining theoretically consistent. Given her explicit distancing of herself from Sartre and even Beauvoir, Butler, I believe, acknowledges this inconsistency.

The argument that Butler should exploit Sartre’s theory is based on the anti-essentialism found in Foucault and Butler. Schrift links this anti-essentialism to Sartre’s denial of a natural essence of the self. As an example of this connection, Schrift points to Butler’s comments on the failure of the term *woman* to define a common identity. If one *is* a woman, that is surely not all one is; the term fails to be exhaustive, not because a pregendered *person* transcends the specific paraphernalia of its gender, but because gender is not always constituted coherently or consistently in different historical contexts, and because gender intersects with racial, class, ethnic, sexual, and regional modalities of discursively constituted identities (ibid).

This comment, according to Schrift, closely resembles Sartre’s claim that the ‘being of being human is to be what it is not and to not be what it is’ (ibid, 16). Of course for Sartre, it is consciousness that makes the being of being human what it is not and not what it is. In the passage quoted above, Butler is not referring to consciousness at all, but to identity. For Butler, identity is nothing other than a sedimentation of performative acts. Identity is created through sustained social performances compelled by compulsory social norms. Rather than an a priori cause, the performative subject is only an effect of repeated iterations of identity performances that have been sedimented over time and practice. Butler’s theory of performative identity is meant to show that there is no prediscursive identity behind the sedimentation of repetitive performances. Moreover, Butler renders the subject identical to performative identity as an effect of rule bound discursive norms. ‘[F]or such appearances are rule-generated entities, ones which rely on the consistent and repeated invocation of rules that condition and restrict culturally intelligible practices of identity. Indeed, to understand identity as a *practice*, and as a signifying practice, is to understand culturally intelligible subjects as the resulting effects of a rule-bound discourse that inserts itself in the pervasive and mundane signifying acts of linguistic life’ (Butler, 1990a, 144-145). Because there are many different compulsory norms that regulate identity categories such as race

and ethnicity, and these intersect with gender identity positions differently for different individuals, there is no one identity that corresponds to the category term *woman*. Identity is multiple and fragmented.

Despite any existentialist residues in their respective works, Foucault and Butler want to distance themselves from Sartre, and particularly from the subject and its agency that they find in his work. Sartre's agency is a function of consciousness' ability to create meaning through its fundamental action of nihilating negation. For Sartre, the for-itself, that is the being of consciousness, confers meaning onto the world in which it finds itself. Sartre does not propose a strong theory of free will, or that consciousness is absolutely free in the sense that choices and ends arise out of nowhere. Consciousness is always in a concrete situation, and choices are limited by the situation. However, Sartre does argue that the for-itself gives meaning to the world and its objects. Thus, the subject, as a for-itself, is constitutive of meaning. Sartre retains the subject as more than simply an effect of discourse. This is what Foucault and Butler find most problematic with Sartre.

Foucault makes the following statement with regard to Sartre's subject centered philosophy: '[Sartre] placed the bare event before or to the side of meaning --the rock of facticity, the mute inertia of occurrence--and then submitted it to the active process of meaning, to its digging or elaboration' (Levy, 2001, 21). Further distancing himself from Sartre, Foucault states that '[t]he point of rupture is situated the day when Levi-Strauss, for societies, and Lacan, for the unconscious, showed us that *meaning* was probably only a sort of surface effect, a shimmering, a foam, and that what traverses us profoundly, what precedes us, sustains us in time and in space, was the *system*' (ibid). Foucault's point is that the subject does not give meaning to itself and its situation by virtue of the for-itself. Rather, as Levi-Strauss and Lacan have shown, meaning precedes the subject. And more forcefully, according to Foucault, Lacan has shown that it is not the subject that speaks, but rather it is the 'structures, the very system of language --and not the subject --that speaks' (ibid). Thus, for Foucault, the subject does not use language to speak the meaning she intends. Rather, language speaks the subject, and the subject is an effect of the system's meaning. The subject is nothing but a function of the system. In direct opposition to this view, Sartre states that 'people have made of speech a language which speaks all by itself. This is an error which should not be made with regard to speech or any other technique...[Otherwise] we shall have lost forever the possibility of meeting the technician' (Sartre, 1993, 662).

Butler makes a move similar to that made by Foucault in order to distance herself from Sartre. She states that her claim that there is no doer behind the deed should not be seen as a return to an 'existential theory of the self as constituted through its acts for the existential theory maintains ...[that] if the subject is culturally constructed, it is nevertheless vested with an agency, usually figured as the capacity for reflexive mediation' (Butler 1990a, 142). It is important to note Butler's rejection of the Sartrean view that the subject has

the capacity for reflexive mediation that vests the subject with agency. To the contrary, Butler insists that the subject is *not* the locus of agency. Butler's model of subjectivity forecloses the possibility of any distancing from the constitutive discursive field (ibid., 145). This view is made clear by Butler with her claim that 'there is no subject prior to its [discursive] constructions' (Butler, 1993, 124).

Butler reduces the subject to a place holder within and effect of discursive power relations, and purges the subject of any robust notion of intention. Butler claims that 'when words engage actions or constitute themselves a kind of action, they do this not because they reflect the power of an individual's will or intention, but because they draw upon and reengage conventions which have gained their power precisely through a *sedimented iterability*. The category of intention, indeed, the notion of "the doer" will have its place, but this place will no longer be behind the deed as its enabling source' (Butler, 1995b). Therefore, in Butler's schema, the *doer* does no work in constructing or even in working through discursive strategies. The enabling conditions for an assertion, the failures of such assertions, and any other displays of agency are precisely and only due to the instability of the system.

Schrift acknowledges Butler's explicit rejection of Sartre's ontology, along with the Sartrean idea of one's life as a project, which requires that one's actions be guided by freely chosen ends or purposes. Schrift, however, does not see this rejection as a major departure from Sartre's theory. Schrift's reading is that Butler 'wants to argue that by adopting the language of "performativity" rather than "expression," the language of Foucauldian "strategies" rather than Sartrean "projects," she is resisting any tendency toward positing some underlying substantive agent' (Schrift, 2001, 15).

I agree that Butler wants to resist positing any underlying substantive agent. In the next passage, however, Schrift trades on an ambiguity between a substantive *agent*, and a notion of substantive *agency* when he writes '[b]ut surely Sartre had no such underlying substantive principle of agency behind his notion of the self as a project of being. Isn't this precisely what is at issue when he demands that we replace the traditional metaphysical claim "I *have* free will" with its existential-ontological reformulation "I *am* free"?' This ambiguity, I believe, is behind Schrift's comparison of Butler and Sartre, and his claim that 'a Sartrean account of subjectivity can be used to further support Butler's...' (ibid). For Sartre, there *is* a subjectivity based in the pre-reflective cogito that *is* free. It maintains its distinction and distancing from constitutive discursive fields and social constructions, and it necessitates that subjects are able to critically reflect upon such constitutive fields. The pre-reflective cogito prevents the subject from collapsing into an effect of discursive relations or constructions. Butler clearly rejects this piece of Sartrean ontology.

Butler's Model of Subjectivity

In her influential work, *Gender Trouble*, Butler (1990) argued for a performative model of subjectivity. Using gender to illustrate this, she argued that identity positions are the reiterations of normative imperatives. Identity positions are the result of performing compulsory social norms. In that text she highlighted the subversive potential of performing gender differently, which implied that subjects could choose which practices they cited or repeated with an eye to directing the consequences of those acts. For example, Butler argued for the subversive potential of drag performances. This conclusion found in *Gender Trouble* was criticized for its argument that the citational slippage inherent in all gender practices could be directed toward a calculable displacement of prevailing sexual norms--in other words, for the contradiction between its non-volitional semiotics and its voluntarist politics (Rothenberg, 1997, 296). This invited the criticism that the performative model of subjectivity was voluntaristic. Butler's most sustained response to the charge of voluntarism is that her theory is decidedly unvoluntaristic and is not a theory of volitional politics. This performativity is not a performance in the sense that an individual decides what gender identity one will be or which gender mask one will wear, and then brings that gender into being by intentionally donning the mask. 'To the extent that a performative appears to "express" a prior intention, a doer behind the deed, that prior agency is only legible *as the effect* of that utterance. For a performative to work, it must draw upon and recite a set of linguistic conventions which have traditionally worked to bind or engage certain kinds of effects' (Butler, 1995b, 134). Butler's most sustained argument (Butler 1993, 1995a, 1995b) 1997) is that through performative practices, discourses produce the subject positions available, and subjects are nothing other than the discursive practices which maintain themselves, through compulsive repetition, as these subject positions.

Given that these practices are generated by the discursive field itself, and that citations are not chosen, Butler also invited the criticism that her model of subjectivity is deterministic. Subject positions are determined in the sense that they are nothing other than the effect of a constitutive system. Butler countered this criticism by arguing for an inherent instability and indeterminacy within discursive fields. Linguistic/discursive determinism is avoided given the variations of repetition that are possible due to the failure of guaranteed outcomes. This unreliability is an inherent feature of Butler's definition of performatives. Throughout her work, Butler has consistently argued that the subject is an effect of discursive systems, but that agency is to be located in the moments of indeterminacy within discursive systems. Whether it is the temporal gap between the performative citations and their effects, the interpellative call and the constituted subject, or the construction of the subject through subjection, in all cases indeterminacy, or the failure of signification, is agency.

In *The Psychic Life of Power*, Butler (1997b) follows Althusser, by arguing that subjects are constituted by language through interpellation. Through interpellation, subjects are called into being by certain forms of address. ‘In the infamous example that Althusser offers, a policeman hails a passerby on the street, and the passerby turns and recognizes himself as the one who is hailed. In the exchange by which that recognition is proffered and accepted, interpellation--the discursive production of the social subject--takes place’ (Butler, 1997b, 5). Butler argues that the constitution of the addressee is not determined by the meaning of the locution. The meaning, the intent can fail. For example, rather than being hurt by being called a queer, one might find it humorous, or feel empowered by the appellation. In short, Butler claims that the meaning can always be resignified.

In her book *Excitable Speech* Butler, following Austin, makes a distinction between locutions that are illocutionary and those that are perlocutionary. Illocutionary acts do in fact perform the deed simultaneously with the utterance, as in the case of *I promise*. They are the performances of an act of saying something. Perlocutionary acts are only instrumental to the accomplishment of actions, but not themselves the accomplishment of the action. They are the effect that a locution may have upon the listener. Thus, with perlocutionary acts there is a gap between the utterance and the effect of the utterance. This gap allows for the possibility of the resignification of meaning and the failure of interpellation to completely constitute the subject (Butler, 1997a, 15). This gap of the perlocutionary act allows for a counter meaning or a talking back that would otherwise be foreclosed.

The distinction between the illocutionary and perlocutionary is necessary to Butler’s criticism of the attempts to censor hate speech as injurious acts.¹ To claim that hate speech constitutes its addressee at the moment of its utterance would be to return to a sovereign subject who holds the power to injure through its speech act. Thus, this rendition of hate speech reinstates the self adequate subject of philosophical discourse which ‘purports to say all that it means, and never to mean more than it actually says’ (Butler, 1987, 196-197). Because Butler thinks this is impossible, there must be a gap between the act and its effect which allows for the failure of a full constitution of the addressee. The gap afforded by perlocutionary acts, according to Butler, allows for the possibility of counter meaning and resignification. All speech acts are vulnerable to failure due to the indeterminacy of meaning within the discursive field.

Since Butler sees the subject as a ‘linguistic category, a placeholder [and a] structure in formation’ (ibid., 27) if an addressee were to resignify the meaning of the derogatory term, that act would be a function of localized discursive relations in which that addressee is situated. This failure of signification allows only for a counter signification, a reversal if you will of the meaning of the term as it appears in the utterance. For example, if one takes a derogatory term to be empowering rather than injurious then the act failed to secure the

¹Some proponents of censorship that Butler examines are Catharine McKinnon and Richard Delgado.

intended consequence. This power of recitation is not a 'function of an individual's *intention*, but is an effect of historically sedimented linguistic conventions...' (Butler 1995b, 134). Given her claim that there is always the possibility of a failure of interpellation to completely constitute the addressee, Butler is not a determinist. She argues that the subject is wholly an effect of social constructions, but that it is not determined by those constructions (Butler, 1993, 124) because of the inherent instability and indeterminacy within the discursive field..

Rather than locating agency in the subject, Butler argues that because the subject is a discursive signification, and 'all signification takes place within the orbit of the compulsion to repeat; *agency* then is to be located in the possible variation on that repetition' (Butler, 1995b, 145). '[T]he subject is a reworking of the very discursive processes by which it is worked.' In this claim we can see the passivity assigned to the *subject*. The subject does not work through its discursive positioning. Rather *it is worked*. Subjects are the 'resulting effects of a rule-bound discourse that inserts itself in the pervasive and mundane signifying acts of linguistic life' (ibid.). The process of signification contains agency within itself in so far as the system is an open system of signs, and as 'historically specific organizations of language, discourses present themselves in the plural, coexisting within temporal frames, and instituting unpredictable and inadvertent convergences from which specific modalities of discursive possibilities are engendered' (ibid.). Critics, however, could still argue that "subjects" are denied any agency in that they are simply functions of the system, indeterminate or not. The discursive constructs, completely, the subject. Although Butler focuses on the failures performatives, citations and interpellations, and argues that no outcomes are ever guaranteed and that all citations can be resignified, she nonetheless makes that resignification a function of the system and not of the subject. The subject may be the site of resignification or redeployment, but only as an effect of such redeployment. Subjects have no capacity to resignify, and therefore, they are still completely the function of the lawful system and its failures.

Butler makes the subject functionally equivalent to Sartre's notion of an empirical self or ego that is constructed by past acts of recitation or repetition. To use Sartrean language, there is no aspect of the for-itself in Butler's model. In Butler's model, subjects are completely mediated by the discursive system with no means of even partial distance or transcendence. Butler's subject is capable only of a misappropriation of the intent of a citation, or a failure to faithfully repeat a normative practice. This forecloses the subject's ability for creative intervention. Clearly, this is not the case for Sartre's subject. Sartre identifies the subject's withdrawal from the world in the act of questioning when he says that the questioner must have the permanent possibility of dissociating himself from the causal order. 'insofar as the questioner must be able to effect in relation to the questioned a kind of nihilating withdrawal, he is not subject to the causal order of the world...Man's relation with being is that he can modify it' (Sartre, 1993, 28-30). It is true that consciousness is always

in situation, however, the choices that arise within and the meaning one gives to that situation are not determined, but are a function of the for-itself and its orientation toward the world. In the mode of being-there, or making meaning out of one's place in the world, Sartre claims that through an act of internal negation I am able 'to escape the "thesis"-- *in the midst of--the world which I am not and by which I make known to myself what I am... [It] is my freedom* which comes to confer on *my* place and to define it as such by situating me' (ibid., 632-633).

According to Sartre, the datum of a situation does not register upon a passive consciousness. Rather, the datum is meaningfully arranged into a situation only because of one's freely chosen ends and purposes. It is those ends and purposes that structure the situation. Sartre states '[f]urthermore this determination of placing, which presupposes all transcendence, can occur only in relation to an end. It is in the light of an end that my place takes on its meaning' (ibid., 633). In order for one to *encounter a situation*, one must do so via the structuring capacity of the for-itself. Without the active conferment of meaning onto a situation, a set of given facts would not constitute an organized situation but only an undifferentiated fullness of being (being in-itself). This *space* for the projection of possibilities, the imagination that things could be different, and the possibility for *making* them different is foreclosed by both Butler and Foucault, as is the active structuring of meaning by consciousness. It is through these actions of consciousness that the meaning of the given world is altered, and emancipatory ends might be pursued.

Conclusion

The reconciliation between Sartre and Butler is dubious. For Sartre, the being of being human is the dialectical tension between being in-itself and being for-itself. To put it in poststructural terms, the subject is part of overarching discursive structures and techniques, but the subject is also able to transcend (at least partially) those structures through the free spontaneity of consciousness. Any situation a subject encounters is, for Sartre, by definition, meaningfully arranged because of the purposes and ends freely chosen by the subject. Butler collapses the subject into identity positions such that they are both the result of discursive relations. For Sartre, the meaning of the self is neither a function of determinacy nor indeterminacy because the subject must transcend the 'facts' of the sedimented self and confer individual meaning on those facts. The Sartrean subject is a mediating subject, one that is always in the midst of culturally regulated norms and resources for signification, but one that is capable of at least partial transcendence.

Bibliography

Boileau, Kevin. (2000). *Genuine Reciprocity and Group Authenticity: Foucault's Developments of Sartre's Social Ontology*. Lanham: University Press of America.

*Sartre and Judith Butler:
A Problem of Agency That Existentialism Can't Solve*

- Butler, Judith. (1987). *Subjects of Desire: Hegelian Reflections in Twentieth-century France*. New York: Columbia University Press.
- _____. (1988.) "Performative Acts and Gender Constitution: An essay in phenomenology and feminist theory" *Theatre Journal* 40.
- _____. (1990a). *Gender Trouble*. New York: Routledge
- _____. (1990b) "Gender Trouble, Feminist Theory and Psychoanalytic Discourse" in Linda Nicholson, ed. *Feminism/Postmodernism*. New York: Routledge.
- _____. (1992). "The Body You Want" *Art forum* 31.
- _____. (1993). *Bodies that Matter*. New York: Routledge.
- _____. (1995a). "Contingent Foundations: Feminism and the question of postmodernism" in Benhabib et al. *Feminist Contentions: A philosophical exchange*. New York: Routledge.
- _____. (1995b). "For a Careful Reading" in Benhabib et al. *Feminist Contentions: A philosophical Exchange*. New York: Routledge.
- _____. (1997) *Excitable Speech: A politics of the performative*. New York: Routledge.
- _____. (1997). *The Psychic Life of Power*. California: Stanford University Press.
- Foucault, Michel. (1984). *The Foucault Reader*. Ed., Paul Rabinow. New York: Pantheon.
- _____. (1980). *Power/Knowledge: Selected interviews and other writings, 1972-1977*. Ed. Colin Gordon. New York: Pantheon.
- Levy, Neil. (2001) "Being Up to Date" *Studies in European Thought* 20.
- Rothenberg, M.A. and Valente, Joseph. (1997). "Performative Chic, the Fantasy of a Performative Politic" *College Literature* 24.
- Sartre, Jean Paul. (1993). *Being and Nothingness*. Trans., Hazel Barnes. New York: Gramercy
- Schrift, Alan. (2001) "Judith Butler: Une Nouvelle Existentialiste?" *Philosophy Today* 45.

CHAPTER TWENTY

The Social Foundations of the Ethical in Kierkegaard's Idea of Existence

Michael J. Matthis

When, in the Concluding Unscientific Postscript, he puts forth the claim that “the only reality that exists for an existing individual is his own ethical reality,”¹ Kierkegaard is not us leading into ethical nihilism, but is developing an ethical position that brings to fruition an ideal of a concrete ethics emphasizing both the autonomy of the subject and the heteronomy of ethical requirement. Such an ideal eludes many philosophies, including Kant's, against which Kierkegaard's ethical ideas compare favorably, as does indeed his idea of community as one that requires the full participation of self in the reality of other self, an ideal that likewise eludes other ethical models of community and selfhood. In order to understand this position, we should observe that this very “ethical reality” of the individual, which Kierkegaard asserts to be the only reality, is one that is unlike the reality of the traditional egoist, or one who, let us say, has immediate and direct knowledge of that reality and essentially no direct knowledge of other realities. Here Descartes comes to mind: all other beings are for him essentially ideas, possibilities, as it were, as the wax in his study, melting in front of him, is not known directly by his senses but only indirectly, in its essence, by his understanding,² and so too concerning other people, whom he “comprehend(s), solely by the faculty of judgment which resides in (his) mind . . .” (M 85), whereas his own reality, that of his thinking self, is so absolutely clear to him that no thing, not even a Super-Thing, such as an evil deceiver, can lead him to question the reality of his own self as a thinking thing (M 93). On this model, to be sure, anyone proposing to develop an ethics should be wary before proceeding. Kierkegaard, however, gives us no indication that his idea of selfhood, in its genuine sense, is even remotely related to Descartes' subjective thinker. What I mean by this is that the self in Descartes' world is in a privileged position with respect to its own reality: each ego *is* its reality as a thinking thing. In this circumstance, Descartes could not fail to be certain of his own existence, since as soon as he becomes clear concerning the truth of the proposition that “there is nothing *easier* for me to understand than my mind”

¹Soren Kierkegaard, Concluding Unscientific Postscript, trans. David F. Swenson and Walter Lowrie (Princeton: Princeton University Press, 1941), 280. Hereafter referred to, in the text, as CUP.

²See Rene Descartes, Meditations, in Discourse on Method and Meditations, trans. Laurence J. Lafleur, Library of Liberal Arts (Indianapolis: The Bobbs-Merrill Co., Inc., 1960, 88. Hereafter referred to, in the text, as M.

(M 91, my italics), then his own existence becomes clear to him as a thing that thinks.¹

Kierkegaard's self, however, is in an entirely different condition with respect to its own existence, and so too in a different condition with respect to other selves. For Kierkegaard existence as such does not bring with it the security of certainty and self-identity but precisely the opposite: existence is always insecure, and the closer the self gets to its own existence, the more the self experiences not certainty but doubt, difficulty, and problems, and any attempt to transform these problems into objects of epistemic certainty serves only to avoid the difficulty and reality of existence. To *exist* is not to *be*, or not *merely* to be, for which reason "precisely because the negative is present in existence, . . . existence is a constant process of becoming . . ." (CUP 75). To exist, then, is not to have the self-identity that is traditionally associated with being, whether of one's self or of any other being, and indeed identity of being characterizes a defection from existence, from the difficulty or process of existing, so that "an existing individual may arrive at identity . . . only by abstracting from existence" (CUP 377). Certainty, the sort that allows Descartes to arrive at immediate knowledge concerning his existence, but also that allows Hegel to construct a system in which all the oppositions of existence melt away by the assistance of abstraction from existence (for which reason "everything said in Hegel's philosophy about process and becoming is illusory" [CUP 272]) belongs to the realm of abstraction, one wherein one is removed from the negativities and difficulties that are intrinsic to existence, and thus Descartes begins his reflections only when "disturbed by no passions . . ." (M 75), and when he has "found a serene retreat in peaceful solitude" (M 75). Descartes' self, that of the *cogito*, does not, for Kierkegaard, actually *exist*, and thus Descartes' certainty is based on a skeptical removal from existence, a certainty based upon an analytical self-containment, as opposed to an involvement in existence as such: ". . . there is an abstract certainty presupposed in all skepticism. . ." (CUP 299n). In the same way, Hegel can construct a system of elaborate logical connections, based on a fundamental skepticism that is empty precisely of existence: ". . . abstract thought gets rid of the difficulty by leaving it out, and then proceeds to boast of having explained everything" (CUP 267-268). Existence, on the other hand, for Kierkegaard, is simply difficulty as such, a difficulty that cannot be captured in or preserved by abstraction, or by epistemological certainty, so that to become realistic with respect to existence is to be *unburdened* of the promise of hope and the possibility of the removal of difficulty:

No, the help in solving the problem of life which Christianity has brought to individual is marked by only one distinguishing feature, namely, the *difficulty*

¹For this reason, concerning Descartes' clarity as to his own mind, we should be cautious before accepting Bernard Williams's assertion that Descartes, in the *Meditations*, in contrast to other of his writings, does not claim to be certain of his thinking, but only of his existence. See *Descartes. The Project of Pure Inquiry* (London: Penguin Books, 1978), 73.

which it has introduced. It is in this fashion that its yoke is easy and its burden light—for him who has thrown all burdens away, those of hope, of despondency, and of despair; but this is a very difficult thing to do (CUP 385, my italics).

Existence, then, is not an abstraction, nor is existence a thing or part of a thing that is to be known; were that the case, difficulty would not be internal to existence but would be resolved in the knowing of existence. Instead of being extrinsic to existence, however, difficulty is inseparable from existence, or “the difficulty again is absolute” (CUP 385), an absoluteness that cannot be avoided by any excuse, such as that existence is *too* difficult, that one cannot be expected to arrive at existence, given the extraordinary problem of so arriving at it, as such a position again removes difficulty from existence as such, making the process one thing, the existent another, and making existence subject to the dialectic of hope and despair, reducing the difficult to an avoidable burden as opposed to an unavoidable reality. As such, however, “existence is . . . a constant striving” (CUP 84), not simply the goal or termination of striving, for which reason existence is *absolutely* difficult, there being no resolution to its difficulty, but therefore never *too* difficult, as the judgment that something is too difficult suggests that the difficult is not absolute but relative, a matter of ease for some, but of a burden for others. Thus the difficulty that is existence is “not subject to a comparative dialectic . . . (but) the difficulty is proportioned absolutely to each individual separately, absolutely requiring his absolute exertion, but no more” (CUP 385). The difficulty of existence, its unresolvability, consists therefore in the fact that existence is not a result or consequence of action, or of traditional *being*, separate from and superior to *becoming*, but just is action as such, the result of which no excuse concerning lack of ability, or bad fortune, or lack of ability, in arriving at the end of action, and an end separate from action itself, is permitted.

The self that we experience, therefore, or the one that we “are,” is one that is precisely not the self of Descartes’ analytically united self, which is why to exist is to become, to act, to confront oneself as a question or problem, and so also to “venture,” to go beyond self-contained identity: “Hence it is not the same individual who makes this venture among others, yielding as a consequence one more predicate to one or another individual” (CUP 379). That is, to venture, or “to become what one already is . . .” (CUP 193), is to disrupt the self as a whole, to replace being with becoming, and in this sense the self becomes “a relation . . . to its own self . . . ,”¹ and so is confronted by itself *as* a question or a problem, a problem that the language of abstraction removes. The self of traditional substance-theory will be capable of epistemological certainty concerning itself, but not of a reflective activity that

¹Soren Kierkegaard, *The Sickness Unto Death*, in *Fear and Trembling and The Sickness Unto Death*, trans. Walter Lowrie (Princeton: Princeton University Press, 1954), 146. Hereafter referred to, in the text, as SUD.

in itself disrupts the being of the self, forcing the self to examine itself, to give an account of itself, placing the self in question to the point of ontological tenuousness, as a result of the “spiritual trial” (JP 1:169) and “daily examination” (JP 1:28) that for Kierkegaard characterize Christianity. To “become subjective,” which Kierkegaard regards as the “highest (task) that is proposed to a human being” (CUP 142), is not to become an impenetrable core of subjectivity, nor a yet more infallible and self-contained individual that not even an evil deceiver can penetrate, but is to become vulnerable, fallible, aware of “the uncertainty of all things (CUP 80), and aware of itself precisely in that uncertainty. Far from isolating the self within a prison of self-identity, along with the illusion of security, and infinite freedom, within that prison, Kierkegaard’s purpose is fundamentally to *limit* the self, to make it the opposite of the immediate self and its wish for unlimited power, but rather to find freedom *in* powerlessness, authority (in a moral sense), and vulnerability: “. . . the religious individual is confined . . . lies fettered in the finite with the absolute conception of God present to him in his human frailty . . .” (CUP 432). In this “frailty” then the self breaks free of its thing-like self-identity and so becomes a *relation*, one wherein the self *must* relate to itself precisely *as* “changed,” that is, as an *other* subject, an other that is, at that very point, *in* the world of selves, and of choices and decisions. “Becoming subjective,” which Kierkegaard also describes as “the most difficult of all tasks . . .” (CUP 146), is therefore fundamentally a becoming objective toward the otherness and difficulty of one’s subjectivity, for which reason Kierkegaard characterizes his idea of subjectivity as a “turning of the subjectivity-principle . . .,” in such a way that “inwardness is shown to be objectivity . . .” (JP 4: 352). The turning of the subjectivity-principle then requires a fundamental orientation of the subject toward itself objectively in such a way that it must make *choices* and *decisions* concerning *itself*, precisely the opposite of the arbitrary, as opposed to choices concerning only others, which means that ethical choice is not a strategy concerning what to do with, how to use, or how to get around an alien other, but is a choice concerning the fundamental equality of selves. Thus “humanness (*Menneskelighed*) . . . is human equality (*Menneske-Lighed*),” whereas “inequality is the inhuman” (JP 1: 24). Kierkegaard is under no illusions concerning the unnaturalness of the transition from ego-centered subjectivity to one that is other-centered: “. . . egotism (*Egoisme*) is one thing, and I-ness or subjectivity (*Egoitet*) is another . . .” (JP 1: 215), and a self that “relate(s) objectively to (its) subjectivity . . .” (JP 4: 348) does not arrive at this orientation without cost to its natural propensity for security: “Christianity is the greatest, most intense, the most powerful restlessness imaginable; it disturbs human existence at its deepest level. . . . It explodes everything, bursts everything” (JP 4: 316).

The self that encounters the problem of itself in existence is therefore already encountering an other, its reality as an eternal problem, and here the eternal itself is not a timeless, “worldless” reality, but is inserted paradoxically within the world as futurity, the locus of decision, a matter of continual difficulty: “When I put eternity and *becoming* together I do not get rest, but

coming into being and futurity” (CUP 272). Existence, then, even as eternal, is created in action, or is a matter of becoming, of “futurity,” rather than its being a thing that I have, something *in* me that I can know concerning myself but not concerning an other’s self. Indeed, what is crucial here for Kierkegaard is that existence, precisely as a matter of action, is not that which isolates me from all others, as occurs in substance versions of selfhood, but is the self’s first real experience of otherness, of the questionable or problematical. The other then is not a reality that the self cannot be or have, owing to the principle of identity, but rather otherness just is existence as that which, in its eternal form, calls upon the self to act from the temporal situation of futurity, and here the self’s “reality” is not merely that of the private ego, known in the immediacy of its own self-identity, but is the self-as-other, as that which the self must become in “becoming subjective.” The self then that cannot just *be* itself, in self-identity, is one that is always a problem, a “synthesis of the temporal and the eternal . . .” (CUP 54 [cf. SUD 146]), a synthesis that, unlike that of Hegel’s, does not take the self conceptually outside of time but rather magnifies time *as* futurity as opposed to abstraction.¹

Abstraction, however, does occur in Kierkegaard, and to the extent that cognition as such is possible, abstraction must occur, but always at the expense of any genuine sense of otherness. What we traditionally posit as the social other is indeed *knowable*, in some sense, for Kierkegaard. However, in its very “knowability,” the other self, to the extent that it is not involved in my *action*, is not truly *other*, which *means* that the other self is not truly *existent*; the existent, that is, is in essence problematical, not just my self’s existence, but any self’s, not a thing to be known, but an other requiring my action, which is why at the ethical or ethical-religious level in Kierkegaard we encounter the equality of all selves. In his second Critique, Kant asserts that the objective reality of objects referred to by ideas such as those of God, freedom, and immortality, is not to be found in the speculative use of reason, but we “can attribute this reality to them in respect to the objects of pure practical reason,”² and so he asserts that with respect to matters of ethical worth, “If something is to be . . . absolutely good or evil. . . , it could not be a thing but only a manner of acting”³ Indeed the real itself, the existent, is a “manner” of action, not of speculation, as “reality is supplied by pure practical reason”⁴ In this same vein, Kierkegaard’s problematical origin of self involves the primacy of the practical and indeterminate over the speculative or theoretical, which means the primacy of the problematical, the “unscientific,” over the cognitive and determinate. Kierkegaard’s version of selfhood then requires that reality be encountered *as* other, even and precisely the reality of oneself as a reality that

¹On the unresolvability of Kierkegaard’s idea of “synthesis,” as opposed to Hegel’s idea of “mediation,” see *Postscript* 75, 268, and 375.

²Immanuel Kant, *Critique of Practical Reason*, trans. Lewis White Beck (Indianapolis: The Liberal Arts Press, 1958), 5.

³*Ibid.*, 62.

⁴*Ibid.*, 141; cf. *Ibid.*, 70: “A practical rule of pure reason, as *practical*, concerns the existence of an object”

is other, and so too as a matter of action, so that choice concerns one's own self as an equal among selves, and as a reality that can and must be acted upon; "knowledge," on the other hand, omits the "difficulty" of existence, its status as a problem for action, so that at the cognitive level the other as such is no longer a problem, meaning no longer a reality, but then not genuinely an other. Hence when he says that "to every other reality he [the self] stands in a cognitive relation . . ." (CUP 380), Kierkegaard is not contending that there actually *is* a social other, the reality of which, as in the tradition, remains distant, other, beyond cognition, in *contrast* to one's own reality; rather he is arguing that since "true knowledge consists of translating the real into the possible" (CUP 380), the other, in its genuine otherness, is no longer in fact a *reality*, since knowledge, at the *cognitive* or speculative level, is not concerned with the other as an autonomous reality, as a problem in relation to which I must indeed become active, and indeed is not concerned with a real other but only with the other as abstract possibility and object of cognition. Only at the ethical level, from the standpoint of action, is reality more than a *mere* possibility, more than an object of cognition, but a difficulty that must be acted upon and that calls forth one's action.

The "possible," then, as object of cognition, is not a problematical other, since only *existence* is the problematical, and thus at no point is Kierkegaard calling for a reduction of reality to (mere) possibility, or the replacement of action with knowledge, since on the contrary action is "the transition from possibility to actuality" (CUP 306), a transition from the *standpoint* of cognition, that which is concerned with the possible, to the *standpoint* of action, or the ethical, that which is concerned with the real, as "reality is the interest of action in existence" (CUP 304). In this case Kierkegaard is not contending that action involves the realization of the possible, as occurs in traditional models of action, but action is concerned with the real itself, *as* a problem, and thus involves a transition away from a cognitive relation to the possible *to* a practical or ethical relation to the real. Just as Kant maintains that the concern of the ethical, and so of action, is in that of the real, so too Kierkegaard places the ethical, or ethical-religious, directly in the real, as a matter of action, so that action is of intrinsic importance, and not merely of extrinsic importance as a means to the development of the real. Kierkegaard rejects the view that an other, in its *reality*, can be an object of a *cognitive* relation, since an *ethical* relation, one that involves the standpoint of action, is required for any meaningful contact with the real or the existent. In so doing, Kierkegaard is demanding a participation by the self or subject *in* the ethical reality of the other (and all) selves, a participation that, in principle, is not possible under traditional models of selfhood that address the other as a metaphysical, as opposed to ethical, reality, one that, in its self-containment, can only be the object of a cognitive relation, through the intercession of ideas, as opposed to an other that, in its very reality, must become the concern of a self's action.

By no means, however, does Kierkegaard remove possibility *from* the real, as possibility is essential if the real is to be a *problem*, an object of *action* in

itself. Kierkegaard does, on the other hand, remove from the standpoint of action possibility as *abstract* possibility, as separate from a determinate, merely complete reality, one that stands opposed to the subject merely as a thing-to-be-known, and reduced to abstract possibility, as opposed to a problem and demand for action. As such possibility that is divorced from existence becomes an object of mere cognition, as possibility converts into a kind of possibility-as-known, one that can be contemplated aesthetically, but as such one that fails to exert a demand upon the knower to act, as possibility must do in order that it can be related to the subject as a *real* or concrete possibility, or as futurity. Kant, for example, in his third Critique, will emphasize the universality of beauty, and therefore its formal character, in opposition to the specific things and objects—including the good—that require a constituting concept of reason or understanding. The beautiful, on the other hand, stands on the periphery of cognition, corresponding to the free play of imagination and understanding as cognitive faculties, and in this free play the universal communicability of beauty, its freedom from existence, becomes its constituting ground. In this regard, however, the agent, “. . . indifferent as regards the existence of the object . . . ,”¹ that is, the object to be judged purely with respect to its form, cannot be *compelled* to recognize its beauty. The judgment of taste for Kant is aesthetic, not logical, and the beautiful, even with its sense of universality, relates directly, not to the object of perception, but to the subject’s delight in the perception itself, even if that delight is imputed to all other subjects, and the subject, as well, is indifferent to its own existence. Thus while Kant will regard the aesthetic, and its formalism, as a transition to the ethical, in insofar as it allows each subject, in its delight in perception, to put itself “in the place of any other man . . . ,”² the absence of compulsion, in the formality of the aesthetic, and in its indifference to existence, is never fully resolved at Kant’s practical level, owing to the essential identity of moral law that determines moral decision and the agent that creates for himself and all others that very law: that is, how can the agent be compelled by the very law that agent creates? Kierkegaard, on the other hand, proceeds to a religious level in order to gain entry into the ethical, and it is at the religious level that the autonomy of subjectivity is itself constituted, without compromising that autonomy, by an other, by the Absolute Paradox, that which constitutes the autonomy of the subject paradoxically by compelling its very free assent to that other. At the concrete level of ethics, therefore, the subject can and must create itself *as* other, and not merely contemplate the universal aesthetically, by concretely placing itself into the reality of the other as an existential demand for action. To this end, Kierkegaard observes that *real*, or concrete, possibility, as opposed to abstract possibility, is central to what he calls “existential communication,” or that which has “the form of a possibility, precisely in order that it may have a relation to existence” (CUP 320). To communicate existence, in this sense,

¹Immanuel Kant, Critique of Judgment, trans. J.H. Bernard (New York: The Free Press, 1951), 43-44.

²Ibid., 136.

does not involve a determinate existence, that which is merely there, as self-contained reality, but involves existence-as-problem, one that communicates a sense of the need for action, and that “compels the recipient to face the problem of existing . . .”(CUP 320). Here then Kierkegaard’s concern is with a possibility that, in its concrete form, involves action, for which reason possibility, in the form of ethical requirement, involves existence as such, so that the self is *free* to choose, and yet *compelled* to choose, as “there must be *no choice*, even though there is a choice” (JP 2: 68, his italics). In this sense requirement offers a choice, an either-or, and at the same time removes choice, *compels* or demands action, whereas a *merely* free will, one that exists abstractly in a kind of timeless zone of deliberation without compulsion, is one that Kierkegaard describes as a “chimera” (JP 2: 67), a “phantom” (JP2: 69), or one that is not engaged in the paradoxical nature of requirement; a *genuine* free-will, on the other hand, is one “that rushes with infinite speed” (JP 2: 68) to the real as such, a reality that exists “in the form of possibility [wherein] it becomes a requirement” (CUP 320).

So too in arguing that ethical action, that which originates in the sense of compulsion that occurs “by apprehending the reality of the other as possibility” (CUP 321), Kierkegaard is not contending that the social other is apprehended as a *mere* possibility, as an *abstract* possibility, any more than God is apprehended as an abstract possibility, precisely because *mere* possibility, as an object of cognition, is not truly *other*, or genuinely a problem that is involved in and compels action. Mere possibility is not dialectical enough to exert a demand upon and compel action, while at the same time allowing action to be free, and in this sense mere possibility represents the sort of non-temporal, reversible character of deliberative thought that is distinct from action: “The relationship of the individual to the action represented in his thought, is still merely a possibility, subject to repudiation” (CUP 304). When Kierkegaard speaks, however, of apprehending the social other’s reality as possibility, he is not attempting to reduce the other’s reality to an *abstract* or *mere* possibility; rather he is contending that the reality of the other needs to be regarded as a possibility with which I have to come to terms: “Everyone who makes a communication, in so far as he becomes conscious of this fact, will therefore be careful to give his existential communication the form of a possibility, precisely in order that it may have a relation to existence” (CUP 320). In this case, Kierkegaard’s social other is a concrete possibility, that which, however, is paradoxical, as the possible, in its very otherness, is also real, underscoring an actuality that gives it a sense of heteronomy, while preserving a sense of possibility that gives the self a sense of autonomy, and thus it is “in the form of possibility,” that the ethical “becomes a requirement” (CUP 320). Those concrete possibilities therefore that serve as both the source of requirement and of freedom, of heteronomy and of autonomy, with respect to the self’s interest in its own reality, are born of limitation as such, the “absurd” or the Absolute Paradox, that which crushes the self (see JP 2: 462), compelling it to obey, and withdraws from the self, requiring freedom (see JP 2: 62), that serves as a continual barrier against a cognitive standpoint that is directed toward abstract,

and non-problematical, possibilities; thus the *esse* of a person cannot be reduced to *posse* (see CUP 285), that which is *merely* thinkable, and yet the relationship of self to the absurd, the God-Man, is itself a “possibility-relationship which every individual has to God” (CUP 139), a possibility-relationship that is concrete because of its essential inscrutability, its origin in the impossible that protects the otherness, the indeterminacy, of the possible as a concrete demand for action, and not an abstract possibility of mere cognition.

The social foundations of the ethical in Kierkegaard therefore have a religious origin, wherein the subject is the recipient of what Kierkegaard calls an “existence-communication” (JP 1: 463). That Kierkegaard’s thought is not burdened by an individuality that is isolated from the reality of other individuals becomes clear when we grasp the problematical, and so “existential,” nature of Kierkegaard’s idea of existence itself, one that he develops initially in the Postscript, and which evolves ultimately into the conception of existence-communication in the Works of Love as one that is concerned with the relation of “man-God-man . . . ,” wherein God is the “middle term”¹ in the relation between self and other self, or in the equality of all selves. That such a position on existential communication does not do violence to his passion for the “the individual,” the category that he regards as his primary contribution toward ethical theory,² Kierkegaard himself denies, insofar as individuality is not a private possession, but “‘the individual’ in its highest measure is beyond a man’s power.”³ As such individuality itself, that is to say, existence, must be communicated in the possibility-relation of self to God, and in the possibility-relation of self to other self, with God, the absolute other and problem as such, as middle term, a communication that Kierkegaard regards as essentially consistent with his idea of the individual’s concern with the problematical reality of the self itself, a reality that is always “other,” a “neighbor” (*neahgebor*) to itself in a position of equality with all selves. Kierkegaard’s ethical individual, sustained and developed first of all within the religion of paradox, is a *significant* and so *singular* self *because* of its own otherness to itself, a self that must examine itself continuously *as* a question and a problem, and as an other in relation to others, a position to which other ethical theories have aspired, but which, remaining within the requirements of logical principles, in place of the requirements of ethical principles, first of all, and adhering as well to substance theories in metaphysics, have been unable to achieve. Kierkegaard corrects this lacuna in the history of ethical theory and in so doing deserves, I believe, our recognition and gratitude, in place of the critical condemnation that he unfortunately has received from some quarters, a debt that, I hope, this study has helped in part to repay.

¹Soren Kierkegaard, The Works of Love, trans. Howard and Edna Hong, Harper Torchbooks (New York: Harper & Row, 1962), 112-113.

²See Soren Kierkegaard, The Point of View for My Work as an Author, trans. Walter Lowrie, Harper Torchbooks (New York: Harper & Row, 1962), 128.

³*Ibid.*, 128.

Bibliography

- Descartes, Rene. Meditations, in Discourse on Method and Meditations. Translated by Laurence J. Lafleur. Library of Liberal Arts. Indianapolis: The Bobbs-Merrill Co., 1960.
- Kant, Immanuel. Critique of Judgment. Translated by J. H. Bernard. New York: The Free Press, 1951.
- _____. Critique of Practical Reason. Translated by Lewis White Beck. Indianapolis: The Liberal Arts Press, 1958.
- Kierkegaard, Soren. Concluding Unscientific Postscript. Translated by David F. Swenson and Walter Lowrie. Princeton: Princeton University Press, 1941.
- _____. Journals and Papers. Translated by Howard and Edna Hong. Four volumes. Bloomington: Indiana University Press, 1967-1975.
- _____. The Point of View for My Work as an Author. Translated by Walter Lowrie, Harper Torchbooks. New York: Harper & Row, 1962.
- _____. Works of Love. Translated by Howard and Edna Hong. Harper Torchbooks. New York: Harper & Row, 1962.
- Williams, Bernard. Descartes: The Project of Pure Inquiry. London: Penguin Books, 1978.

CHAPTER TWENTY ONE

Responsibility in an Age of Global Climate Change: Lessons from the Phenomenology of Thomas Langan

Ingrid Leman Stefanovic

In his insightful book, *A Moral Climate: The Ethics of Global Warming*, Michael Northcott reminds us of the vulnerable times in which we find ourselves today. Citing the conclusions of such venerable international institutions as the Intergovernmental Panel on Climate Change (IPCC), he refers to the growing scientific consensus that the additional 2 degrees Celsius temperature rise conservatively expected by the end of this century is “unprecedented in at least the last 650,000 years of planetary history.” (Northcott, 2007, 20) Increased drought, flooding, famine, human displacement and natural disasters are all linked by experts to a warmer atmosphere which contains 33% more carbon dioxide than in the mid 18th century and which threatens the health of the world as we know it. (Northcott, 2007, 22)

Northcott concludes that “at the heart of the pathology of ecological crisis is the refusal of modern humans to see themselves as creatures, contingently embedded in networks of relationships with other creatures, and with the Creator.” (Northcott, 2007, 16)

This paper enlarges upon this claim, arguing that the phenomenology of Canadian philosopher, Thomas Langan can help to illumine the meaning of these networks of relationships – particularly in light of the moral responsibility that must necessarily be seen to accompany those relationships in an era of global climate change. Building upon the thought of German phenomenologist, Martin Heidegger, Langan has written extensively about how contemporary society has appropriated diverse historical traditions in the search for the meaning of Being. (Langan, 1991; 1996; 2000; 2009) By drawing on one smaller volume in particular, it will be my aim to begin to show how his work also sheds important light on the ontological significance of environmental responsibility in an era of uniquely human-induced global climate change.

Selfhood and Responsibility

One of the most significant contributions of Continental thought to western philosophy has involved its engagement in a critical re-assessment of the nature of the self. No longer pictured atomistically as a solipsistic thinking *thing*, being human has come to be understood in terms of relational contexts

and an understanding of how we are, to use Heidegger's expression, ontologically in-the-world. Heidegger's own phenomenological description of selfhood explicitly avoids a subjectivistic grounding, focusing instead on the fact that human beings (in the German, *Dasein*, or being-there) cannot exist in the absence of a world within which we are situated. The nature of this integral, definitive relation between self and world is not an intellectual construct but is that which is prethematically, spontaneously understood as one grabs a hammer and puts it to use, not as a thing-in-itself separate from me but as part of my action of hammering. "The kind of dealing which is closest to us is...not a bare perceptual cognition but rather that kind of concern which manipulates things and puts them to use; and this has its own kind of 'knowledge.'" (Heidegger, 1962, p.95). For Heidegger, to understand selfhood as *Dasein* is to acknowledge that "its primary kind of Being is such that it is always 'outside' alongside entities which it encounters and which belong to a world already discovered." (Heidegger, 1962, p. 89)

On such a phenomenological reading, we do not discover ourselves as solipsistic individuals first, engaging in relations only secondarily through praxis or by way of abstract speculation. On the contrary, a phenomenological description of the fundamental existential structure of human selfhood reveals that to *be* is to necessarily always also be situated within a specific context and amongst people with whom we co-exist. In short, inasmuch as we *are*, our very being is defined in terms of relationality. As Joan Stambaugh put it, "what is new about this understanding of identity as a relation is that the relation first determines the manner of being of what is to be related and the how of this relation...Identity is belonging-together." (Stambaugh, 1969)

American phenomenologist Edward Casey makes a similar point when he reminds us that "to be" means to *be somewhere*, to be "implaced." (Casey, 1993) In that sense, the "environment" is neither something *out there*, nor is it a *container* of human activities but rather, inasmuch as we exist, we *are* defined in terms of the environment in which we are implaced. We cannot *be* independently of the *environs* that sustains us and in that sense, implacement defines part of the ontological structure of being human. (Stefanovic, 2000)

Tom Langan's work similarly acknowledges and respects such a relational interpretation of selfhood. He notes that, ironically, the more one focuses on the interiority of self, as if the meaning of an isolated identity were to be discovered simply by way of introspection, the more one avoids an entry into authentic selfhood, phenomenologically understood, as temporally and spatially implaced in-the-world. In the words of his small volume, *Self-Discovery*, "authentic discovery of self is the realization of selflessness." (Langan, 1985, 1) To engage in a search of one's selfhood through a psychologistic type of introspection does not take us far, because "genuine discovery of the self is the antithesis of the narcissistic." (Langan, 1985: 1)

On this reading, global climate change is not simply a phenomenon "out there," independent of one's interpretation of selfhood. On the contrary, who we are today is defined, in large part, by the global climate changes that are affecting the world in which we are ontologically implaced. In fact, embedded

in that relational notion of selfhood is, according to Langan, an ethical imperative to act responsibly. Being-in-the-world, then, reveals not simply the ontological structure of selfhood as relational but it reveals a moral dimension as well. “Because we are responsible for one another, because what each of us becomes affects the other, maturation to a full sense of responsibility is at the very heart of the process of becoming fully an *autos*.” (1985, 1) To mature as a self, then, is to actually assume a direct, engaged and embodied responsibility for the world around us and the way in which it is developing.

What, then, might such being “responsible” for the global crisis around environment entail, if we examine the question more closely? Langan suggests that responsibility is tied to caring for another and that such care “can range from taking care of an animal when it manifests hunger (*besorgen*); through care for a friend in distress who needs a manifestation of concern and help out of a dilemma (*Fursorgen*); to care for the communities into which I am thrown and even concern for the world system.” (1985, 49)

In today’s era of global climate change, being responsible certainly entails all three moments described by Langan. Caring for animals is undoubtedly a growing challenge as biodiversity is threatened through species extinction; as factory farming draws animals into the most abominable conditions of living and dying; and as the ecosystems which sustain animal communities are increasingly degraded.

Caring for friends in distress also becomes a special challenge in a world of increasing disparity between developed and developing countries. With much of the burden of environmental devastation falling upon the poor, huge challenges of social justice emerge at this time as well.

To be sure, both of these issues deserve our attention. However, given the planetary scope of the issue and its urgency in an era of global climate change, Langan’s last point – the “concern for the world system” – addresses us in a particularly pressing way. In what sense can we interpret a personal responsibility for the sustainability of our planet in terms that reflect the intimate implacement of selfhood within such a global context?

Langan suggests that, in each instance, my responsibility involves three factors: 1. the *ability* to respond; 2. *Uebereinstimmung* – cor-respondence – with that to which we are called to respond; 3. the *re-pondere* itself – the pledging (*spondeo*) or devotion of ourselves to that which solicits us. (Langan, 1985, 49)

Let us consider each factor in turn, by focusing Langan’s comments upon the issue of global climate change and the issue of sustainability.

The First of Three Factors: The Ability to Respond

Langan sensibly maintains that, if we are to be charged with responsibility for our actions, we must at least have the *ability* to act. There are, however, at least three ways in which our ability to respond to the global climate change

crisis can be impeded. To quote Langan, “(a) the call never reaches us.... (b) we are afraid to respond.... (c) pathological damage makes it impossible for the *Dasein* to hold open the space within which certain hard realities can stand; he flees from them into fantasy.” (1985, 50).

How may the “call never reach us?” Today, whether it is through Al Gore’s film *An Inconvenient Truth* or the fourth IPCC Assessment Report published in 2007, or whether it is through courses taught within formal school programs, the concept of climate change is hardly a secret. (IPCC 2007) At the same time, our lives may be so busy with other pressures, from getting our children to school on time to career demands or other similar, determinate constraints of daily life that “global climate change” might simply seem too abstract a notion to impact upon us in a significant way. Indeed, the diverse activities that fill our days pull us in so many different directions that, if we think of it at all, we may find that focusing on “environment” becomes something else for which we simply do not have sufficient time or energy.

There may also be an issue of what Langan calls one’s “natural faith.” “There is,” as he points out, “little point for the nihilist in talking about a plan for one’s life.” (Langan, 1985, 15) Similarly, there sometimes seems to be little point to expect the SUV driver, self-centredly and needlessly idling his vehicle at the train station, to think in terms of the global commons.

A point to consider is whether the call to responsibility for sustainable action can ever reach such a person and if so, how can such a call be enabled? This question is hardly a trivial one. There are attempts these days by experts in what is called “community-based social marketing” to design prompts and other sorts of devices to incite behavioural changes amongst such people who are oblivious to the environmental impacts of their decisions. (McKenzie-Mohr *et.al.*, 1999) However, larger ethical questions arise regarding the ethics of such behaviour manipulation and, perhaps equally significantly, one wonders about the ability to affect foundational ontological worldviews by what may amount to no more than tinkering with incidental behavioural changes. Addressing the larger question of how to actually influence fundamental paradigms and habitually unsustainable modes of behavior is of increasing concern. (McKenzie-Mohn *et.al.*, 1999)

Another reason to explain why we may be unable to respond to the global environmental crisis is, to quote Langan once again, that “we are afraid to respond.” (1985, 50) Consider the way in which political leaders often pit “environment” in opposition to “economy,” implying that environmental responsibility will simply mean higher taxes and economic loss. Despite the fact that there are multiple examples and huge resources to drawn from to show how environmental sustainability can mean greater financial prosperity, the public often fears the consequences of making changes that will enable improved environmental health because they are not convinced that the financial impact will be bearable. (Labatt and White, 2002)

Finally, Langan suggests that a pathology forces some people to “flee...into fantasy.” (1985, 50) The recent crisis on Wall Street is cited by some as an example where bankers and investors lost touch with the common

good, the meaning of natural capital and the implications of private gain in their fantastical, quantificational manipulations of “phantom wealth.” (Korten, 2009; Patterson, 2010) On another level, simplistic corporate caricatures of environmental activists as “tree-huggers” who stand in the way of progress is another example of a flight into fantasy. It often takes an incident, such as a child coming close to death because of a smog-induced asthma attack, for us to recognize that the environmental crisis is more than an abstract slogan but instead, that it will impact upon our and our children’s very quality of life and well-being.

The Second of Three Factors: Uebereinstimmung – Cor-respondence – with that to which we are called to respond

Recall that to be responsible also means, according to Langan, that a” cor-respondence is invoked with that to which we are called to respond.” (1985, 49) In his book on *Self-Discovery*, we read that “the listening – which is, after all, an activity – must be in profound resonance with the vibrations being sent out by the *Seienden*. In conversation, this means risking hearing the threatening. In observing, I must open myself to the demands of the thing, or the situation, letting the things themselves suggest even the question.” (1985, 51)

Langan is on to something here. The city that arises organically is distinguished from the inhumane, manicured, engineered suburb precisely in terms of how well it allows the “things themselves” to emerge – that is, in terms of the “fitting placement” of the patterns of the wind, the sun, the landscape, the prethematic human experience. (Mugerauer, 1994) Appropriate technologies are precisely those which do not impose themselves upon the environment, but rather, preserve a fitting response to finite, environmental limits. To be responsible, in this context, means to be open to a genuine and humble listening to environmental constraints and opportunities, within which our ways of being can sustainably find their place. (Stefanovic, 2000)

It is worth noting that, contrary to environmental fear-mongers, to engage in environmental care is not simply to pull back, to focus on the finitude of resources, to deny ourselves opportunities. Certainly, the current rate of consumption in the developed world is often callous and steeped in greed and results in consequences that are unsustainable. However, one needs to recognize that a careful listening, in Langan’s sense, invites both humility but also an openness to positive possibilities as well. As Northcott reminds us, “resituating the human economy in the enfolding and relational ecology of the earth is therefore not a dispiriting task of merely constraining or limiting human making and creativity. On the contrary, it is joyous and spiritual work. And the root of this work in those who have already begun it is a renewed appreciation of...the rich history of more beneficent forms of human interaction with life in all its diversity.” (Northcott, 2007, 16)

It is precisely by remaining, in Langan's words, "open to the demands of the thing" in a fitting response to a world in which we stand in relation, rather than responding to the self-centred wants of an egotistic consumer – that the promise of authentic sustainability can find its ontological roots.

The Third Factor: The Re-pondere itself – the Pledging (*spondeo*)

Finally, the third element of responsibility, according to Langan, is the *re-pondere* itself – the "determination to hold open the space by keeping the question open" as well as committing to getting "my hands dirty with the ontical, factual data which the things themselves suggest is needed if progress is to be made." (1985, 51)¹ Langan is calling for a practical commitment of responsible action, which presents challenges to all academics, including philosophers themselves. Perhaps the very first requirement is that those theoreticians who acknowledge their responsibility in principle to advancing global sustainability and planetary health also commit to interdisciplinary dialogue with practitioners in a significantly new way.

In this regard, there is interesting work underway in what Robert Frodeman calls "humanities policy." Recognizing the broad acceptance of the link between science and public policy, he challenges the humanities (including philosophers) to forge a new initiative to impact much more significantly upon the development of environmental policy at local, national and international levels. (Frodeman, 2006) The task that he describes is no small challenge. First, it requires that philosophers themselves are *able* to respond, which means that they must educate themselves in environmental matters. Second, the task requires that they recognize that to genuinely cor-respond to the demands of these environmental challenges, they must be willing to hear voices from other disciplines as well, from the sciences to the social sciences to the arts.

Finally, the task requires that their resolute commitment to "getting their hands dirty with the ontical" will also mean that they engage in forging new public policy by ensuring that their ethical, epistemological and ontological reflections may be seen to be meaningful within the development of such policy initiatives. Translating the philosophical writings into action will require that the discipline of philosophy itself be reconceived. Socrates brought philosophy to the marketplace; in many ways, we must find new ways to take forward precisely such a practice.

Conclusion

Coming to terms with the notion of selfhood as relational in phenomenological terms means that the very structure of one's identity is

¹Heidegger distinguishes between the "ontological" enquiry into the question of the meaning of Being, and "ontic" enquiry, which focuses on beings as discrete entities. See *Being and Time*, Introduction.

intimately tied to the “world” in which it is ontologically implaced. In that vein, the “environment” is not some-thing “out there” but is constitutive of the cultural, geographical and ontological meaning of one’s “self.”

More than this, however, such an ontological structure is never static but, as temporal, the self is called to respond to its being-in-the-world through responsible action. Such responsibility itself, while compelling, is not necessarily without its obstacles. Nevertheless, both the recognition of selfhood as relational, as well as defining that relationality in moral terms, are the conditions of responsible engagement around the current global climate change crisis. If Northcott is right, and the heart of the ecological crisis requires a re-envisioning of selfhood as embedded in networks of relationships, then the phenomenological understanding of being human as ontologically “implaced” becomes all the more compelling. (Northcott, 2007, 16)

In his magnum opus, *Being and Truth*, Tom Langan recognizes the enormity of this task of rethinking selfhood in terms of its relation to global pressures and changes. He reminds us that the “exploitation of our position in the organic hierarchy has allowed our species to extend its territory to the entire planet, and we are now ‘domesticating’ the atmosphere, which bears our aircraft; the deep seas, which harbour our territory-protecting submarines; and even the exosphere, where we orbit our satellites.” (Langan, 1996, 168) Such global influence suggests a human “superiority”, in Langan’s words, by which he means that we are responsible for ensuring that we do not act indiscriminately but that we appreciate the enormity of the challenge that lies ahead in repairing the environmental devastation that we have created.

That challenge invites us to consider how we remain responsible not only for ourselves; not only for other people around us; not only for broader communities; not only for the well-being of other organisms; but also, for the health of our planet. Heidegger once said that “only a God can save us.” He may be right, but we cannot escape the fact that it is also responsible action by responsible human beings, who understand themselves as inextricably integrated with their world rather than isolated from it, that offer the true promise of a healthier planet.

References

- Casey, Edward, *Getting Back into Place: Toward a Renewed Understanding of the Place-World*, (Bloomington, Indiana: Indiana University Press, 1993.)
- Frodeman, Robert, “The Policy Turn in Environmental Philosophy,” in *Environmental Ethics*, Vol. 28, Spring 2006, pp. 3-20.
- Heidegger, Martin, *Being and Time*, trans. By John Macquarrie and Edward Robinson, (New York and Evanston, Harper & Row Publishers, 1962.)
- International Panel on Climate Change, 2007, <http://www.ipcc.ch/ipccreports/index.htm>.
- Korten, David C., *Agenda for a New Economy: From Phantom Wealth to Real Wealth*, (San Francisco: Berrett-Koehler Publishers, Inc., 2009.)

An Anthology of Philosophical Studies

- Labatt, Sonia and Rodney White, *Environmental Finance: A Guide to Environmental Risk Assessment and Financial Products*, (London: Wiley, 2002.)
- Langan, Thomas, *Being and Truth*, (Columbia and London: University of Missouri Press, 1996.)
- , *Human Being: A Philosophical Anthropology*, (Missouri: University of Missouri Press, 2009.)
- , *Self-Discovery*, (San Francisco, California: Golden Phoenix Press, 1985).
- , *Surviving the Age of Virtual Reality*, (Missouri: University of Missouri Press, 2000.)
- , *Truth and Authenticity in the Search for Ecumenic Wisdom* (Missouri: University of Missouri Press, 1991.)
- McKenzie-Mohr and William Smith, *Fostering Sustainable Behaviour: An Introduction to Community-Based Social Marketing*, (British Columbia: New Society Publishers, 1999.)
- Mugerauer, Robert, *Interpretations on Behalf of Place*, (Albany, NY: State University of New York Press, 1994.)
- Northcott, Michael S. *A Moral Climate: The Ethics of Global Warming*, (Maryknoll, New York: Orbis Books, 2007.)
- Patterson, Scott, *The Quants: How a New Breed of Math Whizzes Conquered Wall Street and Nearly Destroyed it*, (New York: Random House, 2010.)
- Stambaugh, Joan, "Introduction" to Martin Heidegger's *Identity and Difference*, (New York, Evanston and London: Harper & Row, Publishers, 1969.)
- Stefanovic, Ingrid Leman, *Safeguarding Our Common Future: Rethinking Sustainable Development*, (Albany, NY: State University of New York Press, 2000.)

CHAPTER TWENTY TWO

From Rationality to Morality: The Challenge of Egoism

James P. Sterba

In Plato's *The Republic* (ca 380BCE), a story is told about a shepherd named Gyges in the service of the king of Lydia, who happened upon a ring that when worn and turned in one direction made him invisible and when turned back made him visible again. Once Gyges became aware of the power of the ring, he arranged to be a messenger to the king. On arriving at the palace, he committed adultery with the queen, attacked the king with her help, killed him, and took over the kingdom.¹

From Plato to the present, this story of Gyges has helped to dramatically pose the question: Why should I be just or moral when I could really benefit from doing something else? More specifically, why should I help people in serious need when I can get a lot of enjoyment out of using my extra income to dine in a fancy restaurant or to take a Mediterranean cruise? Or why should I not cheat on my taxes when I can get away with it? Or why should I forgo any use of resources now so that future generations can have a decent life? This is the fundamental challenge that egoism raises to morality.

Now the Egoist would like to show that the furthering of her own interests is the thing to do. Nevertheless, she recognizes that any reason that she could give for furthering her own interests would suggest a similar or analogous reason that others could give for furthering their interests. Let's imagine that the Egoist, confronts this problem by granting that each person has similar reasons for favoring her own interests. In order to give normative priority to her own interests, The egoist realizes that he must allow that others can and should give the same normative priority to their own interests. It is this willingness to generalize that renders egoism, Universal Ethical Egoism, in which form it constitutes the most serious challenge to morality. Let us now consider three of the most important attempts to meet that challenge.

Appealing to Publicity

Universal Ethical Egoism has been forcefully criticized by contemporary philosopher, Christine Korsgaard, among others, for failing to meet a publicity requirement that is satisfied by morality.² Those committed to morality, just

¹Plato. *The Republic*, trans G.M.A. Grube, (Indianapolis, Hackett, 1974) Book II.

²Although Korsgaard puts her publicity objection to egoism somewhat differently than I have here - she compares egoism to a private language - the egoist's response is to both forms of the objection is the same. Christine Korsgaard, "The Sources of Normativity," *The Tanner Lectures on Human Values* (Salt Lake City, University of Utah Press, 1992). Pp. 20-112.

like those committed to obeying the law, usually want their commitment to be publicly known so that they will be better able to resolve conflicts with others who are similarly committed. By contrast, our Universal Ethical Egoist, is usually not going to want her commitment to Universal Ethical Egoism to be publicly known. This is because if others know that she is an egoist they will tend to guard themselves against being harmed by her and, as a consequence, she may not be able to benefit herself at their expense to the degree that she would otherwise want to do. Rather, while privately endorsing egoism, the egoist is going to publicly, yet hypocritically, profess a commitment to morality in order to secure for herself the benefit that such a public endorsement of morality provides. Of course, privately the egoist thinks that others, like herself, ought to be committed to Universal Ethical Egoism, although she will never tell them so, except perhaps on those occasions when their interests happen to further her own. For her to reveal her commitment to Universal Ethical Egoism on other occasions would work against her interests. There will be times, of course, when the egoist will think that others ought to interfere with her, yet because she and others are publicly committed to a morality that prohibits and attempts to punish interferences of this sort, The egoist thereby hopes that she will be able to avoid such interference to herself. On other occasions, the egoist will be able to further her overall self-interest by selectively, and usually secretly, interfering with the interests of others in violation of the requirements of morality. This is, of course, just what Gyges was able to do. So while Universal Ethical Egoism is not committed to the same publicity requirement that we find in morality, given its rationale for avoiding that requirement, it is difficult to see how this lack of commitment should count as grounds for rejecting the view. Clearly, keeping their commitment to egoism relatively private was essential to the success of Gyges and other real life egoists like Bernard Madoff.¹

Paralleling Egoism and Racism

James Rachels has offered an argument that he thinks “comes closest to an outright refutation of Ethical Egoism.”² Rachels attempts to defeat the view by paralleling egoism with racism and then showing that they are similarly defective. He argues that just as the racist does not provide a good reason why everyone should support the racist’s preferred racial group, so the egoist does not provide a good reason why everyone should support the egoist’s own interests over everyone else’s.

Unfortunately, although Rachels’s directs his argument against egoism generally, his argument only works against Individual Ethical Egoism. It does

¹It should be clear that both Gyges and Madoff as committed to Universal Ethical Egoism, and not to Individual Ethical Egoism, because their behavior no where suggests that they are the only ones who should be acting egoistically.

²James Rachels, *The Elements of Moral Philosophy* 4th edition (New York: McGraw-Hill, 2003), pp. 88-90.

not also work against Universal Ethical Egoism, the view we are presently considering.¹ This is because only Individual Ethical Egoism wants to justify placing just one person in a special category. Universal Ethical Egoism, by contrast, treats everyone the same, at least to the extent of allowing that everyone is equally justified in pursuing his or her own self-interests. So, while Rachels's argument does succeed in defeating Individual Ethical Egoism, it fails to meet the more serious challenge of Universal Ethical Egoism.

Appealing to Consistency

Still another attempt to meet the challenge of Universal Ethical Egoism, advanced by contemporary philosopher, Kurt Baier, among others, tries to show that the view is fundamentally inconsistent.² For the purpose of evaluating this critique, let's use an example, a modern Gyges, Gary Gyges by name, an otherwise normal human being who, for reasons of personal gain, has embezzled \$10,000,000 while working at People's National Bank and is now taking steps to escape to a South Sea island where he will have the good fortune to live a pleasant life protected by the local authorities and untroubled by any qualms of conscience. Suppose that Hedda Hawkeye, a fellow employee, knows that Gyges has embezzled money from the bank and is now about to escape. Suppose, further, that it is in Hawkeye's overall self-interest to prevent Gyges from escaping with the embezzled money because she will be generously rewarded for doing so by being appointed vice-president of the bank. Given that it is in Gyges's overall self-interest to escape with the embezzled money, it now appears that we can derive a contradiction for universal ethical egoism from the following:

1. Gyges ought to escape with the embezzled money.
2. Hawkeye ought to prevent Gyges from escaping with the embezzled money.
3. By preventing Gyges from escaping with the embezzled money, Hawkeye is preventing Gyges from doing what he ought to do.
4. One ought never to prevent someone from doing what he or she ought to do.
5. Thus, Hawkeye ought not to prevent Gyges from escaping with the embezzled money.

Because premise 2 and conclusion 5 are contradictory, Universal Ethical Egoism appears to be inconsistent.

¹Rachels fails to distinguish between these two forms of egoism, blurring them together. This explains why he doesn't see that his argument against the one form of egoism does not work against the other. But he is not alone in failing to distinguish these two forms of egoism. See also Robert Holmes, *Basic Moral Philosophy* 4th edition (Belmont, CA: Wadsworth Publishing Co, 2006), Chapter 5 and Gordon Graham, *Living the Good Life: An Introduction to Moral Philosophy* (New York: Paragon, 1990) Chapter 1.

²Kurt Baier, *The Moral Point of View* (Ithaca: Cornell University Press, 1958), pp.188-191.

The soundness of this argument depends, however, on premise 4, and Seymour, our Universal Ethical Egoist, believes there are grounds for rejecting this premise. As the egoist understands the “oughts” of Universal Ethical Egoism, she is justified in preventing others from doing what they ought to do in violation of premise (4). This is because the egoist understands them to be analogous to the “oughts” of competitive games which do justify just this sort of behavior.

Consider, for example, how in football a defensive player might think that the opposing team's quarterback ought to pass on third down with five yards to go, while not wanting the quarterback to do so and indeed hoping to foil any such attempt the quarterback makes. Or to use Jesse Kalin's example:

I may see how my chess opponent can put my king in check. This is how he ought to move. But believing that he ought to move his bishop and check my kind does not commit me to wanting him to do that, nor to persuading him to do so. What I ought to do is sit there quietly, hoping he does not move as he ought.¹

The point of these examples is to suggest that a Universal Ethical Egoist may, like a player in a game, judge that others ought to do what is in their overall self-interest while simultaneously attempting to prevent such actions or at least refraining from encouraging them. And this provides grounds for rejecting premise (4) of above argument against Universal Ethical Egoism.

The analogy of competitive games also illustrates the sense in which a Universal Ethical Egoist claims that she herself ought to do what is in her overall self-interest. For just as a player's judgment that she ought to make a particular move is followed, other things being equal, by an attempt to perform the appropriate action, so likewise when a Universal Ethical Egoist judges that she ought to do some particular action, other things being equal, an attempt to perform the appropriate action follows.

In general, defenders of Universal Ethical Egoism stress that because we have little difficulty understanding the implications of the use of "ought" in competitive games, we should also have little difficulty understanding the analogous use "ought" by the Universal Ethical Egoist, which in turns provides grounds for rejecting premise (4) of the argument that was supposed to show that Universal Ethical Egoism was an inconsistent view.²

¹Jesse Kalin, "In Defense of Egoism," in *Morality and Rational Self-Interest*, ed. David Gauthier (Englewood Cliffs, NJ: Prentice-Hall, 1970), pp. 73-74.

²To claim, however, that the "oughts" in competitive games are analogous to the "oughts" of universal ethical egoism does not mean there are no differences between them. Most important, competitive games are governed by moral constraints such that when everyone plays the game properly, there are acceptable moral limits as to what one can do. For example, in football one cannot poison the opposing quarterback in order to win the game. By contrast, when everyone holds self-interested reasons to be supreme, the only limit to what one can do is the point beyond which one ceases to benefit. But this important difference between the "oughts" of universal ethical egoism and the "oughts" found in publicly recognized activities like competitive games does not defeat the appropriateness of the analogy. That the "oughts" found

Is There No Way to Meet the Challenge?

The challenge of Universal Ethical Egoism to morality has proven to be a strong one, as the failure of the previous three arguments to meet that challenge shows. In fact, due the past failures to provide a strong defense of morality over egoism, most moral philosophers today have simply given up hope of providing an argument showing that morality is rationally preferable to egoism.¹ Rather, they seem content to show that morality is simply rationally permissible, which means that egoism is rationally permissible as well. Most contemporary moral philosophers do not think anything more can be established.

While this consensus among moral philosophers today is quite strong, there is some hope, expressed by a few philosophers, that we can do better and actually provide an argument showing that morality is rationally required and not just simply rationally permissible.² Most moral philosophers today would certainly like to have such an argument. So given the importance of the question of whether morality can be shown to be rationally required, let us consider just one more attempt to meet the challenge of Universal Ethical Egoism and show that morality is rationally preferable to it.

From Rationality to Morality

Let us begin by imagining that each of us is capable of entertaining and acting upon both self-interested and moral reasons and that the question we are seeking to answer is what sort of reasons for action it would be rational for us to accept.³ This question is not about what sort of reasons we should publicly affirm, since people will sometimes publicly affirm reasons that are quite different from those they are prepared to act upon. Rather, it is a question about what reasons it would be rational for us to accept at the deepest level - in our heart of hearts - when we are speaking truthfully to ourselves.

in publicly recognized activities are always limited by various moral constraints (what else would get publicly recognized?) does not preclude their being a suggestive model for the unlimited action-guiding character of the "oughts" of Universal Ethical Egoism.

¹John Rawls is typical here, as is Thomas Nagel. See Rawls's *A Theory of Justice* (Cambridge: Harvard University Press, 1971) p.136, and Nagel's *The View From Nowhere* (New York: Oxford University Press), p.200ff.

²See, for example, Christine Korsgaard, *The Sources of Normativity*, (Cambridge, Cambridge University Press, 1996).

³"Ought" presupposes "can" here. So unless people have the capacity to entertain and follow both self-interested and moral reasons for acting, it does not make any sense asking whether they ought or ought not to do so. Moreover, moral reasons here are understood to necessarily include (some) altruistic reasons but not necessarily to exclude (all) self-interested reasons. So the question of whether it would be rational for us to follow self-interested reasons rather than moral reasons should be understood as the question of whether it would be rational for us to follow self-interested reasons exclusively rather than some appropriate set of self-interested reasons and altruistic reasons that constitutes the class of moral reasons.

Of course, there are people who are incapable of acting upon moral reasons. For such people, there is no question about their being required to act morally or altruistically. Yet the interesting philosophical question is not about such people but about people, like ourselves, who are capable of acting morally as well as self-interestedly and are seeking a rational justification for following a particular course of action.

In trying to determine how we should act, let us assume that we would like to be able to construct a good argument favoring morality over egoism. Given that good arguments are nonquestion-begging, they do not assume what they are trying to prove, as for example, we would be doing if we were to argue as follows: "God exists because the Bible is the word of God and the Bible says God exists." Accordingly, would like to construct a good argument that does not beg the question like this with respect to the justification of morality.¹

The question at issue here is what reasons each of us should take as supreme, and this question would be begged against Universal Ethical Egoism (hereafter simply egoism) if we proposed to answer it simply by assuming from the start that moral reasons are the reasons that each of us should take as supreme. But the question would be begged against morality as well if we proposed to answer the question simply by assuming from the start that self-interested reasons are the reasons that each of us should take as supreme. This means, of course, that we cannot answer the question of what reasons we should take as supreme simply by assuming the general principle of egoism: "Each person ought to do what best serves his or her overall self-interest."

We can no more argue for egoism simply by denying the relevance of moral reasons to rational choice than we can argue for altruism simply by denying the relevance of self-interested reasons to rational choice and assuming the following general principle of altruism: "Each person ought to do what best serves the overall interest of others."²

Consequently, in order not to beg the question, we have no other alternative but to grant the *prima facie* relevance of both self-interested and moral reasons to rational choice and then try to determine which reasons we would be rationally required to act upon, all things considered.

Notice that in order not to beg the question, it is necessary to back off both from the general principle of egoism and from the general principle of altruism, thus granting the *prima facie* relevance of both self-interested and moral reasons to rational choice. From this standpoint, it is still an open question, whether either egoism or altruism will be rationally preferable, all things considered.

¹For a discussion of other features of good arguments that fortunately are not at issue with regard to providing a justification of morality over egoism, see Merrie Bergmann, James Moor, Jack Nelson, *The Logic Book* (New York: McGraw-Hill, 2008).

²The altruist is here understood to be the mirror image of the egoist. Whereas the egoist thinks that the interests of others count for them but not for herself except instrumentally, the altruist thinks that her own interests count for others, but not for herself except instrumentally.

In this regard, there are two kinds of cases that must be considered: cases in which there is a conflict between the relevant self-interested and moral reasons, and cases in which there is no such conflict.

It seems obvious that where there is no conflict and both reasons are conclusive reasons of their kind, both reasons should be acted upon. In such contexts, we should do what is favored both by morality and by self-interest.

Consider the following example. Suppose you accepted a job marketing a baby formula in a developing country, where the formula was improperly used, leading to increased infant mortality.¹ Imagine that you could just as well have accepted an equally attractive and rewarding job marketing a similar formula in a developed country, where the misuse does not occur, so that a rational weighing of the relevant self-interested reasons alone would not have favored your acceptance of one of these jobs over the other.² At the same time, there were obviously moral reasons that condemned your acceptance of the first job - reasons that you presumably are or were able to acquire. Moreover, by assumption in this case, the moral reasons do not clash with the relevant self-interested reasons; they simply made a recommendation where the relevant self-interested reasons were silent. Consequently, a rational weighing of all the relevant reasons in this case could not but favor acting in accord with both the relevant self-interested and moral reasons.

Now when we rationally assess the relevant reasons in conflict cases, it is best to cast the conflict not as a conflict between self-interested reasons and moral reasons but instead as a conflict between self-interested reasons and altruistic reasons.³ Viewed in this way, three solutions are possible. First, we could say that self-interested reasons always have priority over conflicting altruistic reasons. Second, we could say, just the opposite, that altruistic reasons always have priority over conflicting self-interested reasons. Third, we could say that some kind of compromise is rationally required. In this compromise, sometimes self-interested reasons would have priority over altruistic reasons, and sometimes altruistic reasons would have priority over self-interested reasons.

Once the conflict is described in this manner, the third solution can be seen to be the one that is rationally required. This is because the first and second solutions give exclusive priority to one class of relevant reasons over the other, and only a question-begging justification can be given for such an exclusive

¹For a discussion of the causal links involved here, see *Marketing and Promotion of Infant Formula in Developing Countries*. Hearing before the Subcommittee of International Economic Policy and Trade of the Committee on Foreign Affairs, U.S. House of Representatives, 1980. See also Maggie McComas et al. *The Dilemma of Third World Nutrition* (1983).

²Assume that both jobs have the same beneficial effects on the interests of others.

³This is because, as we shall see, morality itself already represents a compromise between egoism and altruism. So to ask that moral reasons be weighed against self-interested reasons is, in effect, to count self-interested reasons twice - once in the compromise between egoism and altruism and then again when moral reasons are weighed against self-interested reasons. But to count self-interested reasons twice is clearly to show a rationally inappropriate bias in favor of such reasons.

priority. Only by employing the third solution, and sometimes giving priority to self-interested reasons, and sometimes giving priority to altruistic reasons, can we avoid a question-begging resolution.

For example, suppose that you are in the waste disposal business and you have decided to dispose of toxic wastes in a manner that is cost-efficient for you but predictably causes significant harm to future generations. Imagine that there are alternative methods available for disposing of the waste that are only slightly less cost-efficient and will not cause any significant harm to future generations.¹ In this case, you are to weigh your self-interested reasons favoring the most cost-efficient disposal of the toxic wastes against the relevant altruistic reasons favoring the avoidance of significant harm to future generations. If we suppose that the projected loss of benefit to yourself was ever so slight and the projected harm to future generations was ever so great, then a nonarbitrary compromise between the relevant self-interested and altruistic reasons would have to favor the altruistic reasons in this case. Hence, as judged by a nonquestion-begging standard of rationality, your method of waste disposal was contrary to the relevant reasons.

Notice also that this standard of rationality will not support just any compromise between the relevant self-interested and altruistic reasons. The compromise must be a nonarbitrary one, for otherwise it would beg the question with respect to the opposing egoistic and altruistic perspectives.² Such a compromise would have to respect the rankings of self-interested and altruistic reasons imposed by the egoistic and altruistic perspectives, respectively. Since for each individual there is a separate ranking of that individual's relevant self-interested and altruistic reasons (which will vary, of course, depending on the individual's capabilities and circumstances), we can represent these rankings from the most important reasons to the least important reasons as follows:

Individual A	
Self-Interested Reasons	Altruistic Reasons
1	1
2	2
3	3
•	•
•	•
•	•
N	N

¹Assume that all these methods of waste disposal have roughly the same amount of beneficial effects on the interests of others.

²Notice that by "egoistic perspective" here is meant the view that grants the prima facie relevance of both egoistic and altruistic reasons to rational choice and then tries to argue for the superiority of egoistic reasons. Similarly by "altruistic perspective" is meant the view that grants the prima facie relevance of both egoistic and altruistic reasons to rational choice and then tries to argue for the superiority of altruistic reasons.

Individual B	
Self-Interested Reasons	Altruistic Reasons
1	1
2	2
3	3
•	•
•	•
•	•
N	N

Accordingly, any nonarbitrary compromise among such reasons in seeking not to beg the question against either egoism or altruism will have to give priority to those reasons that rank highest in each category. Failure to give priority to the highest-ranking altruistic or self-interested reasons would, other things being equal, be contrary to reason.

Lifeboat Cases

Of course, there will be cases in which the only way to avoid being required to do what is contrary to your highest-ranking reasons is by requiring someone else to do what is contrary to her highest-ranking reasons. Some of these cases will be "lifeboat cases," as, for example, where you and someone else are stranded in a lifeboat that has only enough resources for one of you to survive. But although such cases are surely difficult to resolve (maybe only a chance mechanism, like flipping a coin, can offer a reasonable resolution), they surely do not reflect the typical conflict between the relevant self-interested and altruistic reasons that we are or were able to acquire. Typically, one or the other of the conflicting reasons will rank significantly higher on its respective scale, thus permitting a clear resolution.

Morality as Compromise

Now we can see how morality can be viewed as just such a nonarbitrary compromise between self-interested and altruistic reasons. First, a certain amount of self-regard is morally required or at least morally acceptable. Where this is the case, high-ranking self-interested reasons have priority over low-ranking altruistic reasons. Second, morality obviously places limits on the extent to which people should pursue their own self-interest. Where this is the case, high-ranking altruistic reasons have priority over low-ranking self-interested reasons. In this way, morality can be seen to be a nonarbitrary compromise between self-interested and altruistic reasons, and the "moral reasons" that constitute that compromise can be seen as having an absolute priority over the self-interested or altruistic reasons that conflict with them.

Yet does Morality as Compromise provide an answer to the egoism as practiced by Gyges in myth and by Madoff in reality? Well, it does provide a good, that is, a nonquestion-begging argument favoring morality over egoism and in this way justifies morality over egoism. Of course, this may not have the hoped-for effect on real-life egoists. They may not care whether there is a good argument or justification for what they are doing or proposing to do. To deal with them, we may have to resort to avoidance or coercion. If we do need to resort to coercion, however, Morality as Compromise can also provide us with a good argument for doing so. What more then could we expect it to do to enable us to face the challenge of egoism?

Of course, exactly how Morality as Compromise is to be implemented in practice needs to be determined. So far developed, it is open to a number of different interpretations. A Utilitarian approach would favor one sort of interpretation of the compromise, a Kantian approach another, and an Aristotelian approach yet another, as we will see in subsequent chapters. So Morality as Compromise is anything but a decision procedure for solving practical moral problems. Nevertheless, however this debate between alternative interpretations is resolved, it is clear that some sort of a compromise view or moral solution is rationally preferable to either egoism or altruism when judged from a nonquestion-begging standpoint. Surely, that should suffice to answer the challenge of egoism.

Bibliography

- Sterba, James P. How to Make People Just (Lanham: Rowman & Littlefield, 1988).
_____. Justice for Here and Now (New York, Cambridge University Press, 1998).
_____. The Triumph of Practice Over Theory in Ethics (New York: Oxford University Press, 2005).
_____. "Competing the Kantian Project: From Rationality to Equality" Presidential Address to the APA (*Proceedings and Addresses of the American Philosophical Association*, Vol. 82, Issue 2, Nov. 2008).

